12-2017

Sieve estimation of time-varying panel data models with latent structures

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Xia WANG

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

DOI: https://doi.org/10.1080/07350015.2017.1340299

Published in Journal of Business and Economic Statistics, Available online 21 December 2017, pp. 1-16
https://doi.org/10.1080/07350015.2017.1340299
Pre-print version

Sieve Estimation of Time-Varying Panel Data Models with Latent Structures*

Liangjun Su[a], Xia Wang[b,c], and Sainan Jin[a]

[a] School of Economics, Singapore Management University

[b] Lingnan (University) College, Sun Yat-Sen University

[c] Fujian Key Laboratory of Statistical Sciences, Xiamen University

May 26, 2017

Abstract

We propose a heterogeneous time-varying panel data model with a latent group structure that allows the coefficients to vary over both individuals and time. We assume that the coefficients change smoothly over time and form different unobserved groups. When treated as smooth functions of time, the individual functional coefficients are heterogeneous across groups but homogeneous within a group. We propose a penalized-sieve-estimation-based classifier-Lasso (C-Lasso) procedure to identify the individuals' membership and to estimate the group-specific functional coefficients in a single step. The classification exhibits the desirable property of uniform consistency. The C-Lasso estimators and their post-Lasso versions achieve the oracle property so that the group-specific functional coefficients can be estimated as well as if the individuals' membership were known. Several extensions are discussed. Simulations demonstrate excellent finite sample performance of the approach in both classification and estimation. We apply our method to study the heterogeneous trending behavior of GDP per capita across 91 countries for the period 1960-2012 and find four latent groups.

*Address Correspondence to: Xia Wang, Lingnan (University) College, Sun Yat-Sen University, Guangzhou 510275, China. Email: wangxia25@mail.sysu.edu.cn.

Key Words: Classifier-Lasso; Functional coefficient; Heterogeneity; Latent structure; Panel data; Penalized sieve estimation; Polynomial splines; Time-varying coefficients

1 Introduction

Longitudinal or panel data sets have become widely available nowadays. Analysis of panel data sets has various advantages over that of pure cross-sectional or time series data sets, among which the most important one is perhaps that the panel data provide researchers a flexible way to model both heterogeneity among cross-sectional units and possible structural changes over time. For example, influenced by preference changes, technological progress, institutional transformation, and economic transition, the functional relationships between economic variables may change over time. For this reason, numerous studies have been devoted to test for structural changes in panel data models; see Han and Park (1989), Bai and Lluís Carrion-I-Silvestre (2009), Bai (2010), Kim (2011), Chen and Huang (2014), Li, Qian, and Su (2016), and Qian and Su (2016), among others. On the other hand, panel data usually cover individual units sampled from different backgrounds and with different individual characteristics so that an abiding feature of the data is its heterogeneity, much of which is simply unobserved. Despite the fact that traditional panel data models frequently assume homogeneous slopes for the ease of estimation and inference, such an assumption has been frequently rejected in empirical studies (e.g., Lee, Pesaran, and Smith 1997; Durlauf, Kourtellos, and Minkin 2001; Juárez and Steel 2010; Su and Chen 2013) and there has been increasing interest in modeling slope heterogeneity in panel data models.

Although individual heterogeneity and structural changes are likely to coexist, existing panel data models only address at most one of these two important features. First, the studies on the panel data models with structural changes can be grouped into two categories, one is to consider abrupt changes and the other is to model smooth changes. For the former approach, see, e.g., Bai (2010), Kim (2011), and Qian and Su (2016). The latter approach is mainly motivated from the time-varying (functional) coefficient model or nonparametric regression model in the time series framework. For example, Li, Chen, and Gao (2011) generalize Cai, Fan, and Yao's (2000) and Cai's (2007) time-varying coefficient model to the panel data framework, and develop a local linear dummy variable approach to estimate the functional coefficients; Robinson (2012) introduces a nonparametric trending model with cross-sectional dependence and estimates the trend by kernel method; Chen, Gao, and Li (2012) extend Robinson's (2012) nonparametric trending model to a semiparametric partially linear panel data model. Nevertheless, all parameters of interest, of finite or infinite dimension, in these models are assumed to be common across all cross-sectional units. Second, econometricians and statisticians have tried to address the potential slope heterogeneity in panel data models for a long time, say, through the random coefficient models in econometrics (e.g., Hsiao 2003, Chapter 6; Hsiao and Pesaran 2008) and the random effects model in statistics (e.g., Diggle, Heagerty, Liang, and Zeger 2003, Chapter 9). More recently, Su, Shi, and Phillips (2016, SSP hereafter) propose a novel variant of Lasso to estimate heterogeneous linear panel data models where the slope parameters are heterogeneous across groups but homogeneous within a group and the group membership is unknown. But they do not allow the coefficients to change over time.

In this paper we propose a heterogeneous time-varying panel data model with latent group structures to capture individual heterogeneity and smooth structural changes over time simultaneously. To the best of our knowledge, this is the first model to capture these two important features together. As individual heterogeneity and smooth structural changes are likely to coexist, our model appears more realistic than existing models and is expected to have much broader empirical applications. Following Cai (2007), we model the time-varying coefficients as smooth functions of time which can be estimated by nonparametric sieve or kernel methods. We could allow each individual unit to have distinct functional coefficients and estimate them individually but only with a slow convergence rate. Here, we adopt the latent group structure advocated by SSP and assume that the individuals belong to K different groups, and the individual functional coefficients are heterogeneous across groups but homogeneous within a group. The major difficulty lies in the fact that the individuals' group membership is unknown. Our interest is to infer the individuals' group membership and estimate the group-specific functional coefficients at the same time.

In terms of statistical methodology, we propose a penalized-sieve-estimation-based classifier-

Lasso (C-Lasso) procedure to identify the individuals' membership and to estimate the group-specific functional coefficients simultaneously. Since our estimation procedure is an iterative procedure and computationally involved, we prefer the sieve method to the kernel method in order to approximate the unknown functional coefficients. In particular, we propose to use polynomial B-splines given their good approximation properties and stable numerical properties; see, e.g., Huang, Wu, and Zhou (2004), Huang and Shen (2004), and Xue and Yang (2006). The penalty term in our penalized sieve estimation (PSE) is constructed in the spirit of SSP's C-Lasso procedure which aims to shrink each individual coefficient to one of the K unknown groups. Our procedure achieves classification and estimation in a single step. The classification exhibits the desirable property of uniform consistency. The PSE-based C-Lasso estimators and their post-Lasso versions achieve the oracle property of Fan and Li (2001) so that the group-specific functional coefficients can be estimated as well as if the individuals' membership were known. We also propose a data-driven method to determine the number of groups. Simulations demonstrate excellent finite-sample performance of our approach in both classification and estimation. We apply our method to study the heterogeneous trending behavior of GDP per capita across 91 countries for the period 1960-2012 and find four latent groups.

It is worth mentioning that recently grouping or homogeneity pursuit has generated a lot of interest in statistics. The fused Lasso of Tibshirani, Saunders, Rosset, Zhu, and Knight (2005) can be regarded as an effort of exploring slope homogeneity. Bondell and Reich (2008) propose a method called OSCAR to simultaneously select variables while grouping them into predictive clusters. Shen and Huang (2010) develop an algorithm called grouping pursuit by using the truncated L_1 penalty to penalize differences for all pairs of coordinates. Such an algorithm is further extended by Zhu, Shen, and Pan (2013) to allow for simultaneous grouping pursuit and feature selection. To explore homogeneity of coefficients, Ke, Fan, and Wu (2015) propose a new method called clustering algorithm in regression via data-driven segmentation (CARDS), which is extended to the panel setup by Wang, Phillips, and Su (2017). Nevertheless, almost all of these papers consider linear data models in the cross-sectional framework.

The rest of the paper is organized as follows. In Section 2, we introduce our time-varying panel data model with latent group structures. In section 3, we consider the PSE for this model. We examine the asymptotic properties of the estimators in Section 4 and discuss several possible extensions in Section 5. Section 6 provides Monte Carlo study and empirical illustration. Section 7 concludes. All proofs of the main results are relegated to Appendix A. Further technical details and the numerical algorithm are contained on the online supplementary appendix.

Notation. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\| (\equiv [\mathrm{tr}(AA')]^{1/2})$ and its Moore-Penrose generalized inverse as A^{+}. When A is symmetric, we use $\mu_{\max}(A)$ and $\mu_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. Let $\|A\|_{\mathrm{sp}} (\equiv [\mu_{\max}(AA')]^{1/2})$ denote the spectral norm of A. $\mathbb{I}_a$ and $\mathbf{0}_{a\times b}$ denote the $a \times a$ identity matrix and $a \times b$ matrix of zeros. $\mathbf{1}\{\cdot\}$ denotes the indicator function. We use "p.s.d.' to abbreviate "positive semidefinite". The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{D}$ convergence in distribution, and plim probability limit. We use $(N, T) \to \infty$ to signify that N and T tend to infinity jointly. For a vector-valued function $\alpha(\cdot)$ defined on $[0,1]$, we use $\|\alpha\|_2$ to denote its L_2-norm: $\|\alpha\|_2 \equiv \{\int_0^1 \|\alpha(v)\|^2 dv\}^{1/2}$. Given sequences of positive numbers a_{NT} and b_{NT}, $a_{NT} \lesssim b_{NT}$ and $b_{NT} \gtrsim a_{NT}$ mean a_{NT}/b_{NT} is bounded, and $a_{NT} \asymp b_{NT}$ means that both $a_{NT} \lesssim b_{NT}$ and $a_{NT} \gtrsim b_{NT}$ hold. When a_{NT} and b_{NT} are random, $a_{NT} \lesssim b_{NT}$ and $b_{NT} \gtrsim a_{NT}$ mean a_{NT}/b_{NT} is stochastically bounded and $a_{NT} \asymp b_{NT}$ means that both a_{NT}/b_{NT} and b_{NT}/a_{NT} are stochastically bounded.

2 Time-Varying Panel Structure Model

In this section, we introduce the time-varying panel structure model. The dependent variable Y_{it} is generated according to the following time-varying panel structure model:

$$Y_{it} = \gamma_i + \beta_{it}' X_{it} + u_{it}, \;\; u_{it} = \sigma_i(X_{it})\,\varepsilon_{it}, \tag{2.1}$$

where $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, X_{it} is a $p \times 1$ vector of regressors, γ_i's are unobserved individual fixed effects that may be correlated with some components of X_{it} and are assumed to be different for different individuals, ε_{it} has mean zero and variance one and is independent of the process $\{X_{it}\}$ so that u_{it} is the idiosyncratic error term with conditional variance

$\sigma_i^2(X_{it})$ given X_{it}, and $\beta_{it} = \beta_i(t/T)$ is a $p \times 1$ vector of time-varying slope coefficients exhibiting the following latent group structure:

$$\beta_{it} = \sum_{k=1}^{K} \alpha_k(t/T) \cdot \mathbf{1}\{i \in G_k\}. \tag{2.2}$$

We assume that $\|\alpha_j - \alpha_k\|_2 \neq 0$ for any $j \neq k$, $\cup_{k=1}^{K} G_k = \{1, 2, \ldots, N\}$, and $G_j \cap G_k = \varnothing$ for any $j \neq k$. Let $N_k = \#G_k$ denote the cardinality of the set G_k. For the moment we assume that the number of groups, K, is known and fixed, but each individual's group membership is unknown. We will propose an information criterion to determine K in Section 4.4.

Interestingly, our model in (2.1) and (2.2) does not appear as restrictive as the time-invariant panel data models considered in Lin and Ng (2012), Bonhomme and Manresa (2015), and SSP. All the latter authors assume that an individual cannot change its group identity during the whole sampling period. As a matter of fact, this restrictive assumption also serves as an important motivation for our paper. To see this point, we can go back to the SSP's framework. When the regression coefficients do not change over time, the model is given by

$$Y_{it} = \gamma_i + \beta_i' X_{it} + u_{it},\ i = 1, ..., N,\ t = 1, ..., T,$$

where β_i's have some grouped patterns. For simplicity, suppose that there are only two groups with the first and second half of individuals belonging to Groups 1 and 2, respectively. In this case, the number of groups (2 here) and the group identity for each individual remain fixed during the whole time period. To allow for the change of group membership for some individuals, it is natural to model β_i as $\beta_i(t/T)$. In this case, we say that individuals i and j belong to the same group (say Group 1) if only if $\beta_i(t/T) = \beta_j(t/T)$ for $t = 1, ..., T$. It is possible that

$$\beta_i(t/T) = \beta_j(t/T) \text{ for all } t = 1, ..., T_0 \text{ and } \beta_i(t/T) \neq \beta_j(t/T) \text{ for some } t = T_0 + 1, ..., T,$$

in which case i and j belong to the same group (say Group 1) until time T_0 and different groups after that. In this case, the total number of groups is generally not 2 but 3 at least, and our PSE method introduced below can identify the emergence of new groups asymptotically. That is, by enlarging the number of groups (K), we effectively allow the change of group membership for some individuals over the whole time period. In essence, the number of groups should not be regarded as given at the beginning of the sampling period. Instead, it is determined throughout the whole sampling period, which makes our model very attractive in comparison with existing panel structure models. In short, the change of group membership has been built into our model through the use of time-varying functional coefficients.

Our interest is to estimate the time-varying group-specific functional coefficients $\alpha_k(\cdot)$, $k = 1, 2, ..., K$, and to infer each individual's group identity. Following the literature on smooth time-varying regression models (e.g., Cai 2007; Robinson 2012; Chen, Gao, and Li 2012, Zhang, Su, and Phillips 2012), we assume that $\beta_i(\cdot)$'s and $\alpha_k(\cdot)$'s are smooth functions of t/T. See also Robinson (1989, 1991) for the discussion on the use of t/T rather than t as an argument of the functions.

Our model (2.1) is fairly general, and it includes a variety of panel data models as special cases.

1. If $X_{it} = 1$ and $\beta_i(\cdot) = \beta(\cdot)$ for some function $\beta(\cdot)$ and for each $i = 1, ..., N$, then the model in (2.1) becomes the nonparametric trending panel data model studied by Robinson (2012):
$$Y_{it} = \gamma_i + \beta(t/T) + u_{it}. \tag{2.3}$$

2. If $\beta_i(\cdot) = \beta(\cdot)$ for some function $\beta(\cdot)$ and for each $i = 1, ..., N$, then (2.1) becomes the time-varying functional coefficient panel data model studied by Li, Chen, and Gao (2011):
$$Y_{it} = \gamma_i + \beta(t/T)' X_{it} + u_{it}. \tag{2.4}$$

3. If $\beta_i(v) = \beta_i$ and $\alpha_k(v) = \alpha_k$ for any $v \in (0, 1]$, $i = 1, ..., N$, and $k = 1, ..., K$, then model (2.1) becomes the linear time-invariant panel structure model considered by SSP.

4. If $X_{it} = 1$, then model (2.1) becomes the nonparametric trending panel structure model:
$$Y_{it} = \gamma_i + \beta_i(t/T) + u_{it}, \tag{2.5}$$
where $\beta_{it} = \beta_i(t/T)$ satisfies the latent group structure in (2.2). Obviously, this model generalizes that of Robinson (2012) to allow for heterogeneous trending behavior.

In sum, our model in (2.1) can be regarded as an extension of that of SSP or that of Li, Chen, and Gao (2011). It extends the time-invariant model of SSP to allow time-varying coefficients and the homogeneous functional coefficient model of Li, Chen, and Gao (2011) to allow heterogeneous time-varying functional coefficients. It captures the smooth structural changes over time and the individuals' heterogeneity across groups simultaneously, and is thus expected to have much broader empirical applications than existing models in the literature. For example, as our empirical application demonstrates, the logarithm of the gross domestic product (GDP) per capita across countries exhibit heterogenous grouped patterns over time. For another example, the beneficial effects of foreign direct investment (FDI) on economic growth in host countries may exhibit both smooth structural changes and cross-country heterogeneity (c.f., Cai, Chen, and Fang 2014). In either case, one has to apply the methodology developed in this paper.

Hereafter, we use the superscript 0 to denote the true values or functions. In particular, we use $\beta_i^0(\cdot)$ and $\alpha_k^0(\cdot)$ to denote the true functional coefficients and G_k^0 the true value of G_k.

3 Penalized Sieve Estimation

In this section, we introduce the PSE method.

3.1 Sieve Approximation of Time-Varying Coefficients

We propose to estimate $\beta_i(v)$ and $\alpha_k(v)$ by polynomial splines of order d. Let $J_0 = J_0(N,T)$ be a prescribed integer that depends on (N,T). Divide $[0,1]$ into (J_0+1) subintervals $I_j = [v_j, v_{j+1})$ for $j = 0, 1, ..., J_0 - 1$ and $I_{J_0} = [v_{J_0}, 1]$, where $\mathcal{V} \equiv \{v_j\}_{j=1}^{J_0}$ is a sequence of equally spaced points (interior knots),

$$v_{-(d-1)} = ... = v_{-1} = v_0 = 0 < v_1 < v_2 < ... < v_{J_0} < 1 = v_{J_0+1} = ... = v_{J_0+d},$$

$v_j = jh$ for $j = 1, ..., J_0$, and $h = 1/(J_0+1)$ denotes the distance between two neighboring points. Let $\mathbb{G} = \mathbb{G}_{d,\mathcal{V}}$ denote the space of polynomial splines of order d based on $\mathcal{V}$. It consists of functions g satisfying: (i) g is a polynomial of degree $d-1$ on each of the subintervals

$\{I_j\}_{j=0}^{J_0}$, (ii) for $d \geq 2$, g is $d-2$ times continuously differentiable on $[0,1]$. Let $J = J_0 + d$. We use $B(v) = (B_{-d+1}(v), B_{-d+2}(v), ..., B_{J_0}(v))'$ to denote a basis system of the space $\mathbb{G}$. In this paper, we focus on B-splines of order d (or degree $d-1$) because of the good approximation properties of splines and the stable numerical properties of B-splines. In particular, we will use cubic B-splines in our simulations and application, corresponding to $d = 4$. For more discussions on splines or B-splines, we refer the readers directly to Schumaker (1981), DeVore and Lorentz (1993), de Boor (2001), or the survey paper by Chen (2007). See Appendix A for some basic properties of B-splines that are used in our analysis.

Given the spline basis system $B(v)$, we can approximate the square-integrable functions $\beta_i(v)$ and $\alpha_k(v)$ by $\pi_i' B(v)$ and $\omega_k' B(v)$ for some $J \times p$ matrices $\pi_i = (\pi_{i,1}, ..., \pi_{i,p})$ and $\omega_k = (\omega_{k,1}, ..., \omega_{k,p})$. Note that for notational simplicity we choose the same basis functions with the same interior knots and polynomial order to approximate different functions of interest. Then we can rewrite the model in (2.1) as:

$$Y_{it} = \gamma_i + [X_{it} \otimes B(t/T)]' \operatorname{vec}(\pi_i) + e_{it} \tag{3.1}$$

where $e_{it} = u_{it} + \beta_{it}' X_{it} - [X_{it} \otimes B(t/T)]' \operatorname{vec}(\pi_i)$, and $\pi_i = \omega_k$ if $i \in G_k$ for $i = 1, ..., N$ and $k = 1, ..., K$.

3.2 Penalized Sieve Estimation of π_i and ω_k

Given the representation of the model in (3.1), we could estimate π_i by minimizing the following least squares objective function:

$$Q_{0,NT}(\boldsymbol{\pi}, \boldsymbol{\gamma}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{Y_{it} - \gamma_i - Z_{it}' \operatorname{vec}(\pi_i)\}^2,$$

where $\boldsymbol{\pi} = (\operatorname{vec}(\pi_1)', ..., \operatorname{vec}(\pi_N)')'$, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)'$, and $Z_{it} \equiv X_{it} \otimes B(t/T)$. Since the individual effects γ_i's are not of primary interest, we concentrate them out and obtain the following concentrated objective function:

$$Q_{1,NT}(\boldsymbol{\pi}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tilde{Y}_{it} - \tilde{Z}_{it}' \operatorname{vec}(\pi_i)\right]^2, \tag{3.2}$$

where $\tilde{Z}_{it} = Z_{it} - \frac{1}{T}\sum_{t=1}^{T} Z_{it}$ and $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$. By minimizing the objective function in (3.2), we obtain the least squares estimator of $\boldsymbol{\pi}$ by $\tilde{\boldsymbol{\pi}} = (\text{vec}(\tilde{\pi}_1)', ..., \text{vec}(\tilde{\pi}_N)')'$, where

$$\text{vec}(\tilde{\pi}_i) = \left(\frac{1}{T}\sum_{t=1}^{T} \tilde{Z}_{it}\tilde{Z}_{it}'\right)^{+} \left(\frac{1}{T}\sum_{t=1}^{T} \tilde{Z}_{it}\tilde{Y}_{it}\right) \text{ for } i = 1, ..., N. \tag{3.3}$$

Let $\boldsymbol{\omega} = (\text{vec}(\omega_1)', \ldots, \text{vec}(\omega_K)')'$ and $\tilde{Z}_i = (\tilde{Z}_{i1}, \ldots, \tilde{Z}_{iT})'$. To estimate $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$ together, we consider the following penalized least squares objective function:

$$Q_{NT,\lambda}^{(K)}(\boldsymbol{\pi}, \boldsymbol{\omega}) = Q_{1,NT}(\boldsymbol{\pi}) + \frac{\lambda}{N}\sum_{i=1}^{N} \tilde{\sigma}_i^{2-K} \prod_{k=1}^{K} \left\| \tilde{V}_i \text{vec}(\pi_i - \omega_k) \right\| \tag{3.4}$$

where $\lambda = \lambda(N, T)$ is a tuning parameter, $\tilde{V}_i = \{\text{diag}(\frac{J}{T}\tilde{Z}_i'\tilde{Z}_i)\}^{1/2}$, and $\tilde{\sigma}_i = \{\frac{1}{T}\sum_{t=1}^{T}[\tilde{Y}_{it} - \tilde{Z}_{it}'\text{vec}(\tilde{\pi}_i)]^2\}^{1/2}$ is an estimator of the sample standard deviation of $\{u_{it}\}_{t=1}^{T}$. Minimizing objective function in (3.4) yields the PSE-based *classifier-Lasso* (C-Lasso hereafter) estimators $\hat{\boldsymbol{\pi}} = (\text{vec}(\hat{\pi}_1)', ..., \text{vec}(\hat{\pi}_N)')'$ and $\hat{\boldsymbol{\omega}} = (\text{vec}(\hat{\omega}_1)', ..., \text{vec}(\hat{\omega}_K)')'$ of $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$, respectively.

The objective function in (3.4) is in the same spirit as that in SSP if we replace $\tilde{\sigma}_i$ and $\tilde{V}_i$ by one and an identity matrix, respectively. We apply $\tilde{\sigma}_i$ and $\tilde{V}_i$ to ensure the scale-invariant property of our objective function: $Q_{NT}(\boldsymbol{\pi}, \boldsymbol{\omega})$ remains unchanged when one changes the scales of either $\tilde{Z}_{it}$ or $\tilde{Y}_{it}$ by changing those of X_{it} and Y_{it}. Note that the objective function in (3.4) is not convex in $\boldsymbol{\pi}$ or $\boldsymbol{\omega}$. In the supplementary Appendix C we provide an iterative algorithm to obtain the estimators $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\omega}}$. Given these estimators, we can obtain the estimators of $\beta_i(v)$'s and $\alpha_k(v)$'s as follows:

$$\hat{\beta}_i(v) = \hat{\pi}_i' B(v), \text{ and } \hat{\alpha}_k(v) = \hat{\omega}_k' B(v) \text{ for } i = 1, ..., N, \text{ and } k = 1, ..., K. \tag{3.5}$$

We will study the asymptotic properties of these estimators in the next section.

Remark 1. Alternatively, one can extend the K-means algorithm to our framework. The latter approach is adopted by Ng and Lin (2012) in linear panel data models with additive fixed effects, by Bonhomme and Manresa (2015) for linear models with grouped additive effects, and by Ando and Bai (2016) in linear panel data models with grouped interactive fixed effects. There are three major differences between the C-Lasso and K-means methods. First, the C-Lasso estimation needs to specify the number of groups (K) and the tuning parameter (λ) while the K-means estimation requires the specification of K only. Despite this, it is hard to tell which method should be preferred as the additional parameter λ may offer some degree of freedom in finite samples. Secondly, the K-means algorithm forces all individuals to be classified into one of the K groups while the C-Lasso procedure may leave some individuals unclassified for small values of λ. For large values of λ, the C-Lasso can also classify all individuals to one of the K groups and produce similar results as the K-means algorithm. But it is hard to tell whether we should force all individuals to be classified. In fact, when T is not large, forcing all individuals to be classified via either the K-means algorithm or the use of a large value of λ for the C-Lasso tends to yield a large proportion of misclassification. In contrast, when λ is not large enough, the C-Lasso allows for some individuals to be left unclassified, which could yield better finite sample performance for the estimators of the group-specific functional coefficients especially when T is not large. For large T, the choice of λ does not matter very much and the two methods generally produce highly consistent classification results. Third, computationally the C-Lasso is much less demanding than the K-means algorithm. This is because the K-means estimation is NP-hard and the C-Lasso problem, despite its nonconvexity, can be transformed into a sequence of convex problems (see the supplementary Appendix C).

We will show that C-Lasso estimators of the group-specific functional coefficients and their post-Lasso versions are oracally efficient – they are asymptotically equivalent to the corresponding infeasible estimators of the group-specific functional coefficients that are obtained by knowing all individual group identities. Following the theoretical studies in Bonhomme and Manresa (2015) and Ando and Bai (2016), we conjecture that the K-means estimators also exhibit the oracle property. If this is the case, the two types of estimators for the group-specific functional coefficients are asymptotically equivalent.

4 Asymptotic Theory

In this section we first establish the preliminary convergence rates for $\hat{\beta}_i(v)$ and $\hat{\alpha}_k(v)$, and then study the consistency of the classification. We also establish the asymptotic distributions of $\hat{\alpha}_k(v)$'s and their post-Lasso versions and study the determination of K.

4.1 Preliminary Rates of Convergence for Coefficient Estimates

Let $\min_{i,t}$ and $\max_{i,t}$ denote $\min_{1\leq i\leq N}\min_{1\leq t\leq T}$ and $\max_{1\leq i\leq N}\max_{1\leq t\leq T}$, respectively. Let $C^{(\gamma)}[0,1]$ denote the space of functions that are γth order continuously differentiable on $[0,1]$, where $\gamma\geq 1$. Let $X_{it}^{(2)}=X_{it}$ if X_{it} does not contain 1 and $X_{it}=(1,X_{it}^{(2)\prime})'$ otherwise.

To study the consistency of $\hat{\beta}_i(v)$ and $\hat{\alpha}_k(v)$, we make the following assumptions.

Assumption A1. (i) Let $X_i^{(2)}=(X_{i1}^{(2)},...,X_{iT}^{(2)})'$ and $\varepsilon_i=(\varepsilon_{i1},...,\varepsilon_{iT})'$. $\{(X_i^{(2)},\varepsilon_i)\}$ are independently distributed over i.

(ii) For each $i=1,...,N$, the process $\{(X_{it}^{(2)},\varepsilon_{it}),\,t=1,2,...\}$ is strong mixing with mixing coefficient $\alpha(j)$ satisfying $\alpha(j)\leq c_\alpha\rho^j$ for some $c_\alpha<\infty$ and $\rho\in[0,1)$.

(iii) $\max_{i,t}E\|X_{it}\|^q\leq\bar{c}_x<\infty$ and $\max_{i,t}E|u_{it}|^q\leq\bar{c}_u<\infty$ for some $q>6$.

(iv) There exist positive constants $\underline{c}_{xx}$ and $\bar{c}_{xx}$ such that $\underline{c}_{xx}\leq\min_{i,t}\mu_{\min}(\mathrm{Var}\,(X_{it}^{(2)}))\leq\max_{i,t}\mu_{\max}(E(X_{it}X_{it}'))\leq\bar{c}_{xx}$ whenever $X_{it}\neq 1$. There exists $\underline{c}_\sigma>0$ such that $\lim_{T\to\infty}\min_i\bar{\sigma}_{i,T}^2\geq\underline{c}_\sigma$, where $\bar{\sigma}_{i,T}^2\equiv T^{-1}\sum_{t=1}^T E(u_{it}^2)$.

(v) For $k=1,2,...,K$, $\alpha_k^0\in C^{(\gamma)}[0,1]$ for some $2\leq\gamma\leq d+1$. There exists $\underline{c}_\alpha>0$ such that $\min_{1\leq j\neq k\leq K}\left\|\alpha_j^0-\alpha_k^0\right\|_2\geq\underline{c}_\alpha$.

(vi) $N_k/N\to\tau_k\in(0,1)$ for each $k=1,...,K$ as $N\to\infty$.

Assumption A2. (i) As $(N,T)\to\infty$, $J\to\infty$, $J^2/T\to 0$, $\lambda J^{(K+1)/2}\to 0$, and $N^2T^{1-q/2}(\ln N)^{q\epsilon_0/2}\to 0$ for some $\epsilon_0>1$.

(ii) As $(N,T)\to\infty$, $J/\ln T\to\infty$, $\lambda J^{\gamma+(K-1)/2}\to\infty$ and $\lambda\sqrt{T}J^{(K-1)/2}/(\ln T)^3\to\infty$, and $\lambda(\ln T)^v\to 0$ for some $v>0$.

Assumptions A1(i)-(ii) require that $\{X_{it}^{(2)},\varepsilon_{it}\}$ be independently distributed over individuals and weakly dependent over time. We assume that $\{(X_{it}^{(2)},\varepsilon_{it}),\,t=1,2,...\}$ is a strong mixing process with a geometric decay rate, which can be satisfied by many well-known linear processes such as ARMA processes and a variety of nonlinear processes. Note that we allow serial correlation in $\{u_{it}\}$ and lagged dependent variables in $X_{it}^{(2)}$. When $X_{it}^{(2)}$ contains lagged dependent variables (e.g., $Y_{i,t-1}$), the strong mixing condition imposes some restrictions on the fixed effects λ_i and the error terms u_{it}. In this case, we can assume that λ_i's are nonrandom and u_{it}'s have Lebesgue-integrable characteristic functions (Andrews 1984). If λ_i's are stochastic, we can follow Hahn and Kuersteiner (2011) and Su and Chen (2013) and adopt the concept of conditional strong mixing where the mixing coefficients are defined by conditioning on the fixed effects. A1(iii) imposes moment conditions for X_{it} and u_{it}. A1(iv) imposes the identification condition that ensures the large dimensional matrix $\frac{J}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$ (see (3.3)) is asymptotically nonsingular and the preliminary estimator $\tilde{\sigma}_i^2$ of $\bar{\sigma}_{i,T}^2$ is uniformly bounded away from zero with probability approaching one (w.p.a.1); see Lemmas A.3 and A.5 in Appendix A. Note that X_{it} may contain 1 or not and we allow $X_{it}=1$. When $X_{it}=1$, the first part of Assumption A1(iv) is not relevant. Assumption A1(vi) is also assumed in SSP and it implies that each group has an asymptotically non-negligible number of members as $N\to\infty$.

The first part of Assumption A1(v) imposes smooth conditions on the group-specific functional coefficients α_k^0 (and thus the individual functional coefficients β_i^0). By Theorem 12.6 in de Boor (2001, p.149), there exists $\omega_k^0\in\mathbb{R}^J$ such that

$$\sup_{v\in[0,1]}\left\|\alpha_k^0(v)-\omega_k^{0\prime}B(v)\right\|=O(h^\gamma)=O\left(J^{-\gamma}\right)\text{ for }k=1,...,K. \tag{4.1}$$

Similarly, there exists $\pi_i^0\in\mathbb{R}^J$ such that

$$\sup_{v\in[0,1]}\left\|\beta_i^0(v)-\pi_i^{0\prime}B(v)\right\|=O(h^\gamma)=O\left(J^{-\gamma}\right)\text{ for }i=1,...,N, \tag{4.2}$$

and $\pi_i^0=\omega_k^0$ if $i\in G_k^0$. The second part of A1(v) implies conditions for the identification of the group-specific functional coefficients. By the triangle inequality, (A.1) in Appendix A, and (4.1), we have

$$\begin{aligned}\underline{c}_\alpha &\leq \left\|\alpha_j^0-\alpha_k^0\right\|_2\leq\left\|\left(\omega_j^0-\omega_k^0\right)'B\right\|_2+\left\|\alpha_j^0-\omega_j^{0\prime}B\right\|_2+\left\|\alpha_k^0-\omega_k^{0\prime}B\right\|_2\\ &= \left\{\operatorname{tr}\left(\left(\omega_j^0-\omega_k^0\right)'\int B(v)\,B(v)'\,dv\left(\omega_j^0-\omega_k^0\right)\right)\right\}^{1/2}+O\left(J^{-\gamma}\right)\\ &\asymp J^{-1/2}\left\|\omega_j^0-\omega_k^0\right\|\text{ for any }j\neq k.\end{aligned}$$

That is,

$$\left\|\omega_j^0-\omega_k^0\right\|\asymp J^{1/2}\text{ for any }j\neq k, \tag{4.3}$$

which will be used in the proof of Theorem 4.1.

Assumptions A2 imposes conditions on N,T, J, and λ. It requires that λ shrinks to zero at a suitable rate such that the penalty term can effectively distinguish individuals in one group from those in the other groups asymptotically. The range in which λ converges to zero mainly depends on T and J but not N. The intuition is clear: J controls the bias from the sieve approximation and the effective number of parameters in the sieve estimation; T, in conjunction with J, controls the speed at which one can estimate the individual functional coefficients $\pi_i(\cdot)$'s and the group-specific functional coefficients $\omega_k(\cdot)$'s. Clearly, A2 allows the choice of a wide range of values of λ and J provided the corresponding functions are sufficiently smooth and q is large enough.

The following theorem studies the preliminary convergence rates of the estimators of π_i^0 and ω_k^0.

Theorem 4.1 *Suppose Assumptions A1 and A2(i) hold. Then*
(i) $\|\hat{\pi}_i - \pi_i^0\| = O_P(J^{-\gamma+1/2} + JT^{-1/2} + \lambda J^{(K+1)/2})$ *for* $i = 1, 2, ..., N$,
(ii) $N^{-1}\sum_{i=1}^N \|\hat{\pi}_i - \pi_i^0\|^2 = O_P(J^{-2\gamma+1} + J^2T^{-1})$,
(iii) $\left\|\hat{\omega}_{(k)} - \omega_k^0\right\| = O_P(J^{-\gamma+1/2} + JT^{-1/2})$ *for* $k = 1, 2, ..., K$,
where $(\hat{\omega}_{(1)}, ..., \hat{\omega}_{(K)})$ *is a suitable permutation of* $(\hat{\omega}_1, ..., \hat{\omega}_K)$.

Theorems 4.1(i) and (ii) establish the pointwise and mean-square convergence of $\hat{\pi}_i$, respectively. The first two terms, namely, $J^{-\gamma+1/2}$ and $JT^{-1/2}$ in part (i) reflect the contributions of the usual asymptotic bias and variance terms of sieve estimation, respectively, and the last term $\lambda J^{(K+1)/2}$ signifies the effect of the penalty term in the C-Lasso procedure. For small enough λ, i.e., if $\lambda \lesssim \max(T^{-1/2}J^{(1-K)/2},\ J^{-\gamma-K/2})$, we obtain the usual convergence rate for the coefficient estimates when B-splines are used. Interestingly, the mean-square convergence of $\hat{\pi}_i$ and the pointwise convergence of $\hat{\omega}_{(k)}$ do not depend on λ, which is analogous to the results of SSP in the parametric setting. See the proof in Appendix A for details. In particular, we show in the proof of Theorem 4.1(iii) that the convergence rate of $\hat{\omega}_{(k)}$ depends on the mean-square but not the pointwise convergence rate of $\hat{\pi}_i$. Note that Assumption A2(i) ensures that $\|\hat{\pi}_i - \pi_i^0\| = o_P(1)$ and $\left\|\hat{\omega}_{(k)} - \omega_k^0\right\| = o_P(1)$.

For notational simplicity, hereafter we will write $\hat{\omega}_k$ for $\hat{\omega}_{(k)}$ and $\hat{\alpha}_k(\cdot)$ for $\hat{\alpha}_{(k)}(\cdot)$ where $\hat{\alpha}_{(k)}(\cdot) = \hat{\omega}_{(k)}' B(\cdot)$. Then we can define the estimated groups:

$$\hat{G}_k = \{i \in \{1, 2, ..., N\} : \hat{\pi}_i = \hat{\omega}_k\} \text{ for } k = 1, ..., K. \tag{4.4}$$

The following corollary establishes the pointwise and L_2 convergence rates of $\hat{\beta}_i(\cdot)$ and $\hat{\alpha}_k(\cdot)$.

Corollary 4.2 *Suppose Assumptions A1 and A2(i) hold. Then*

(i) $\sup_{v\in[0,1]}\left\|\hat{\beta}_i(v)-\beta_i^0(v)\right\|=O_P(J^{-\gamma+1/2}+JT^{-1/2}+\lambda J^{(K+1)/2})$ *and* $\int_0^1\left\|\hat{\beta}_i(v)-\beta_i^0(v)\right\|^2dv$ $=O_P(J^{-2\gamma}+JT^{-1}+\lambda^2J^K)$ *for* $i=1,2,...,N$*;*

(ii) $\sup_{v\in[0,1]}\|\hat{\alpha}_k(v)-\alpha_k^0(v)\|=O_P(J^{-\gamma+1/2}+JT^{-1/2})$ *and* $\int_0^1\|\hat{\alpha}_k(v)-\alpha_k^0(v)\|^2dv=$ $O_P(J^{-2\gamma}+JT^{-1})$ *for* $k=1,2,...,K$*.*

Similar results hold when we replace the integration by the sample mean. That is, $\frac{1}{T}\sum_{t=1}^{T}||\hat{\beta}_i(t/T)-\beta_i^0(t/T)||^2=O_P(J^{-2\gamma}+JT^{-1}+\lambda^2J^K)$ for $i=1,2,...,N$, and $\frac{1}{T}\sum_{t=1}^{T}||\hat{\alpha}_k(t/T)-\alpha_k^0(t/T)||^2=O_P(J^{-2\gamma}+JT^{-1})$ for $k=1,2,...,K$.

4.2 Classification Consistency

We define the following sequences of events:

$$\hat{E}_{k,NT,i}=\left\{i\notin\hat{G}_k|i\in G_k^0\right\}\text{ and }\hat{F}_{k,NT,i}=\left\{i\notin G_k^0|i\in\hat{G}_k\right\},$$

where $i=1,2,...,N$ and $k=1,2,...,K$. Let $\hat{E}_{k,NT}=\cup_{i\in G_k^0}\hat{E}_{k,NT,i}$ and $\hat{F}_{k,NT}=\cup_{i\in\hat{G}_k}\hat{F}_{k,NT,i}$. The events $\hat{E}_{k,NT}$ and $\hat{F}_{k,NT}$ mimic Type I and Type II errors in statistical tests: $\hat{E}_{k,NT}$ denotes the error event of not classifying an individual in the kth group into the k-th group; $\hat{F}_{k,NT}$ denotes the error event of classifying an individual that does not belong to the k-th group into the k-th group. Following SSP's definition, we say that the classification is *uniformly consistent* if $P(\cup_{k=1}^K\hat{E}_{k,NT})\to0$ and $P(\cup_{k=1}^K\hat{F}_{k,NT})\to0$ as $(N,T)\to0$, i.e., the probability of committing either type of errors shrinks to zero asymptotically.

The following theorem establishes the classification consistency of our method.

Theorem 4.3 *Suppose Assumptions A1 and A2 hold. Then*

(i) $P(\cup_{k=1}^K\hat{E}_{k,NT})\leq\sum_{k=1}^K P(\hat{E}_{k,NT})\to0$ *as* $(N,T)\to\infty$*;*

(ii) $P(\cup_{k=1}^K\hat{F}_{k,NT})\leq\sum_{k=1}^K P(\hat{F}_{k,NT})\to0$ *as* $(N,T)\to\infty$*.*

Theorem 4.3 implies that all individuals within a group, say G_k^0, can be simultaneously correctly classified into the same group (denoted $\hat{G}_k$) w.p.a.1. Conversely, all individuals that are classified into the same group, say $\hat{G}_k$, simultaneously correctly belong to the same group (G_k^0) w.p.a.1.

Remark 2. Let $\hat{G}_0$ denote the group of individuals in $\{1, 2, ..., N\}$ that are not classified into any of the K groups, i.e., $\hat{G}_0 = \{1, 2, ..., N\} \backslash (\cup_{k=1}^{K} \hat{G}_k)$. Define the events $\hat{H}_{iNT} = \{i \in \hat{G}_0\}$. Theorem 4.3(i) implies that $P\left(\cup_{1 \le i \le N} \hat{H}_{iNT}\right) \le \sum_{k=1}^{K} P(\hat{E}_{kNT}) \to 0$. That is, all individuals can be classified into one of the K groups w.p.a.1. Nevertheless, when T is not large, it is possible for a small number of individuals to be left unclassified if we stick with the classification rule in (4.4). To ensure that all individuals are classified into one of the K groups, say, if one is sure that there are no isolated individuals as in our simulations, one can modify the classifier a little bit. In particular, for any $i \in \hat{G}_0$ we can classify it to $\hat{G}_l$ for some $l = 1, ..., K$ if

$$\|\hat{\pi}_i - \hat{\omega}_l\| = \min\{\|\hat{\pi}_i - \hat{\omega}_1\|, ..., \|\hat{\pi}_i - \hat{\omega}_K\|\}.$$

Since the event $\cup_{1 \le i \le N} \hat{H}_{iNT}$ occurs with probability approaching zero, we can follow SSP to ignore it in subsequent asymptotic analysis and restrict our attention to the classification rule in (4.4) to avoid confusion. That is, $\hat{G}_k = \{i \in \{1, ..., N\} : \hat{\pi}_i = \hat{\omega}_k\}$ for $k = 1, ..., K$.

4.3 Post-Lasso Estimator and Oracle Property

Given the estimated groups $\{\hat{G}_k, k = 1, ..., K\}$ defined in (4.4), we can readily pool the observations within each estimated group to obtain the post-Lasso sieve estimator of the corresponding group-specific functional coefficients by:

$$\hat{\alpha}_{\hat{G}_k}(v) = \hat{\omega}'_{\hat{G}_k} B(v), \tag{4.5}$$

where for $k = 1, ..., K$,

$$\text{vec}\left(\hat{\omega}_{\hat{G}_k}\right) = \left(\sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}'_{it}\right)^{+} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Y}_{it}. \tag{4.6}$$

When the group identity for each individual is known, we obtain the oracle estimators: $\hat{\alpha}_{G_k^0}(v) = \hat{\omega}'_{G_k^0} B(v)$, where $\text{vec}(\hat{\omega}_{G_k^0}) = \left(\sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}'_{it}\right)^{+} \sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Y}_{it}$.

Let $u_i = (u_{i1}, ..., u_{iT})'$. Then $\text{Var}(u_i|X_i) = \Sigma_i^{1/2} V_i \Sigma_i^{1/2}$, where $\Sigma_i = \text{diag}(\sigma_i^2(X_{i1}), ..., \sigma_i^2(X_{iT}))$ and $V_i = E(\varepsilon_i \varepsilon_i')$. Let $X_{it}^{(\sigma)} = \sigma_i(X_{it}) X_{it}$. We add the following assumption:

Assumption A3. (i) For $k = 1, ..., K$, there exist two positive constants $\underline{c}_V$ and $\bar{c}_V$ such that $0 < \underline{c}_V \le \lim_{(N,T)\to\infty} \min_{i \in G_k^0} \mu_{\min}(V_i) \le \lim_{(N,T)\to\infty} \max_{i \in G_k^0} \mu_{\max}(V_i) \le \bar{c}_V \delta_{NT}$ for some nondecreasing sequence δ_{NT} satisfying $\delta_{NT} N^{-1} \to 0$.

(ii) $\max_{i,t} E\left\|X_{it}^{(\sigma)}\right\|^{q} \leq \bar{c}_x^{(\sigma)}$ for some constant $\bar{c}_x^{(\sigma)} < \infty$ and $q > 6$.

(iii) There exist two positive constants $\underline{c}_{xx}^{(\sigma)}$ and $\bar{c}_{xx}^{(\sigma)}$ such that $\underline{c}_{xx}^{(\sigma)} \leq \min_{i,t} \mu_{\min}(\mathrm{Var}(X_{it}^{(2,\sigma)})) \leq \max_{i,t} \mu_{\max}(E(X_{it}^{(\sigma)} X_{it}^{(\sigma)\prime})) \leq \bar{c}_{xx}^{(\sigma)}$, where $X_{it}^{(2,\sigma)} = X_{it}^{(\sigma)}$ if $X_{it}^{(\sigma)}$ does not contain nonrandom element and $X_{it}^{(2,\sigma)}$ is a collection of the random elements of $X_{it}^{(\sigma)}$ otherwise.

(iv) As $(N,T) \to \infty$, $NTJ^{-2\gamma} \to 0$.

Assumption A3(i) imposes conditions on the variance-covariance matrix of ε_i in order to verify the Lindeberg condition for a central limit theorem to hold. For its minimum eigenvalue, we only need it to be bounded away from zero by a positive constant. Such a condition can be easily satisfied. For example, if we follow Huang, Wu, and Zhou (2004) and assume that ε_{it} can be decomposed into two components: $\varepsilon_{it} = \varepsilon_{it}^{(1)} + \varepsilon_{it}^{(2)}$, where $\varepsilon_{it}^{(1)}$ is an arbitrary mean zero process and $\varepsilon_{it}^{(2)}$ is an independent process of "measurement errors" that are independent over time and have mean zero and constant positive variance σ^2, then the lower bound for the minimum eigenvalue is given by σ^2. For the maximum eigenvalue, we allow it to be constant or divergent as $(N,T) \to \infty$. If there is no serial correlation, then V_i is diagonal and the condition requires that the maximum value of the diagonal elements of V_i to be of order $O(\delta_{NT})$, where $\delta_{NT} = o(N)$. For any $m \times n$ matrix $A = \{a_{ij}\}$, note that $\|A\|_{\mathrm{sp}}^2 \leq \|A\|_1 \|A\|_\infty$, where $\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |a_{ij}|$ and $\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |a_{ij}|$. Since V_i is symmetric and p.s.d., we have $\|V_i\|_1 = \|V_i\|_\infty$ and $\mu_{\max}(V_i) = \|V_i\|_{\mathrm{sp}} \leq \|V_i\|_1$. So our maximum eigenvalue condition will be satisfied if the column sums of V_i are bounded by the order $\delta_{NT} = o(N)$. This condition is automatically satisfied under our strong mixing and moment conditions on $\{\varepsilon_{it}\}$ if ε_{it} has the same second moment across i. Assumption A3(i) says that the central limit theorem can hold without the strong mixing conditions or identical moments across individuals. Assumptions A3(ii) and (iii) parallel the first part of A1(iii) and A1(iv), respectively. If $\sigma_i(X_{it}) = \underline{\sigma}_0 > 0$ a.s., A3(ii) and (iii) would become redundant. A3(iv) ensures that the asymptotic biases of the estimators $\hat{\alpha}_k(v)$ and $\hat{\alpha}_{\hat{G}_k}(v)$ do not contribute to their limit distributions.

The following theorem gives the oracle property of the PSE-based C-Lasso estimators and their post-Lasso versions.

Theorem 4.4 *Suppose Assumptions A1-A3 hold. Then, for $k = 1, 2, ..., K$, we have*

(i) $\sqrt{N_k T/J}\mathbb{S}_k^{-1/2}\left[\hat{\alpha}_k(v) - \alpha_k(v)\right] \xrightarrow{D} N(0, \mathbb{I}_p)$,

(ii) $\sqrt{N_k T/J}\mathbb{S}_k^{-1/2}\left[\hat{\alpha}_{\hat{G}_k}(v)-\alpha_k(v)\right]\xrightarrow{D} N(0,\mathbb{I}_p)$,
where $\mathbb{S}_k^{-1/2}$ *is the symmetric square root of* $\mathbb{S}_k^{-1}$, $\mathbb{S}_k=(\mathbb{I}_p\otimes B(v))'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\{\frac{J}{N_kT}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}(\mathbb{I}_p\otimes B(v))$, *and* $\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}=\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$.

Theorem 4.4 indicates that both the C-Lasso estimator $\hat{\alpha}_k(v)$ and the post-Lasso version $\hat{\alpha}_{\hat{G}_k}(v)$ are asymptotically equivalent to the infeasible estimator $\hat{\alpha}_{G_k^0}(v)$. The latter can be obtained only if one knows all individuals' group identity. In this sense, our C-Lasso estimators and their post-Lasso versions have the oracle efficiency. Despite the asymptotic equivalence between the C-Lasso and post-Lasso estimators, it is well known that the post-Lasso estimators typically outperform the C-Lasso estimators and are thus recommended for practical use.

Remark 3. As a referee points out, it does not appear transparent to see the relative rates on N and T to obtain all the asymptotic properties so far because they are related to Assumptions A2(i)-(ii) and A3(iv). To gain some insight, we focus on the case where all functions of interest are sufficiently smooth so that the approximation biases are asymptotically negligible and all terms associated with γ in Assumptions A2(ii) and A3(iv) do not matter. In this case, the single most important requirement on (N,T) is given by the last part of Assumption A2(i) because other conditions are essentially imposed on J and λ. This part of the assumption holds if q or T or both are sufficiently large. If $\{(X_{it}^{(2)},u_{it})\}$ is sub-Gaussian as in Bonhomme and Manresa (2015), $q=\infty$ and N can grow at any polynomial rate faster than T. The first two conditions in Assumption A2(i) require that J diverge to infinity at a rate slower than $\sqrt{T}$ (i.e., $1\ll J\ll T^{1/2}$) and all the other conditions in Assumption A2 would be satisfied if

$$\max\left(J^{-\gamma-\frac{K-1}{2}},T^{-1/2}J^{-\frac{K-1}{2}}(\ln T)^3\right)\ll\lambda\ll\min\left(J^{-\frac{K+1}{2}},\ J^{-1}/(\ln T)^{\upsilon}\right).$$

It is easy to see that such λ exists under that condition that $1\ll J\ll T^{1/2}$ and $\gamma\geq 2$. When N and T pass to infinity at the same rate (as is commonly assumed in large dimensional macro panels), our choice of J_0 (or $J\equiv J_0+d$) and λ in the following simulation and application sections would meet the above conditions and Assumption A3(iv) provided $q>6$ and $\gamma>3$.

Remark 4. For statistical inference, one needs to estimate $\mathbb{S}_k$ consistently. Suppose that one can estimate $\Theta_k\equiv\frac{J}{N_kT}\sum_{i\in\hat{G}_k}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i$ by $\tilde{\Theta}_k$ such that $||\tilde{\Theta}_k-\Theta_k||_{\mathrm{sp}}=o_P(1)$.

Define

$$\tilde{\mathbb{S}}_k = (\mathbb{I}_p \otimes B(v))' (J\tilde{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1} \tilde{\Theta}_k (J\tilde{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1} (\mathbb{I}_p \otimes B(v)),$$

where $\tilde{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} = \frac{1}{\hat{N}_k T} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}_{it}'$ and $\hat{N}_k = \#\hat{G}_k$. Following SSP we can readily show that $\hat{N}_k / N_k \xrightarrow{P} 1$, $||J\tilde{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} - J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}||_{\text{sp}} = o_P(1)$, and $\tilde{\mathbb{S}}_k - \mathbb{S}_k = o_P(1)$. Under various primitive conditions, one can propose the corresponding consistent estimator $\tilde{\Theta}_k$; see, e.g., Su and Jin (2012). The procedure is standard and thus omitted.

4.4 Determination of the Number of Groups

In practice, the exact number of groups, K, is typically unknown. Here we assume that K is bounded from above by a finite integer $K_{\max}$ and denote the true value of K as K_0. We propose a BIC-type information criterion (IC) to determine the data-driven choice of K.

By minimizing the objective function in (3.4), we obtain the C-Lasso estimates $\{\hat{\pi}_i(K,\lambda)\}_{i=1}^N$ and $\{\hat{\omega}_k(K,\lambda)\}_{k=1}^K$ of $\{\pi_i\}_{i=1}^N$ and $\{\omega_k\}_{k=1}^K$ for $K = 1, ..., K_{\max}$, where we make the dependence of the estimators on (K,λ) explicit. When $K = 1$, one effectively works on the non-penalized sieve estimation. As before, we can classify individual i into group $\hat{G}_k(K,\lambda)$ if and only if $\hat{\pi}_i(K,\lambda) = \hat{\omega}_k(K,\lambda)$, i.e.,

$$\hat{G}_k(K,\lambda) = \{i \in \{1,2,...,N\} : \hat{\pi}_i(K,\lambda) = \hat{\omega}_k(K,\lambda)\} \text{ for } k = 1, ..., K. \tag{4.7}$$

Denote $\hat{G}(K,\lambda) = \{\hat{G}_1(K,\lambda), ..., \hat{G}_K(K,\lambda)\}$. Conditional on the classification, we could define the post-Lasso estimate of ω_k as follows:

$$\text{vec}\left(\hat{\omega}_{\hat{G}_k(K,\lambda)}\right) = \left(\sum_{i \in \hat{G}_k(K,\lambda)} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}_{it}'\right)^{+} \sum_{i \in \hat{G}_k(K,\lambda)} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Y}_{it}. \tag{4.8}$$

In addition, define $\hat{\sigma}^2_{\hat{G}(K,\lambda)} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in \hat{G}_k(K,\lambda)} \sum_{t=1}^{T} \left[\tilde{Y}_{it} - \tilde{Z}_{it}' \text{vec}(\hat{\omega}_{\hat{G}_k(K,\lambda)})\right]^2$. Then, we choose $\hat{K}$ to minimize the following information criterion:

$$IC(K,\lambda) = \ln\left[\hat{\sigma}^2_{\hat{G}(K,\lambda)}\right] + \rho_{NT} JpK \tag{4.9}$$

where ρ_{NT} is a tuning parameter.

Let $G^{(K)} = (G_{K,1}, ..., G_{K,K})$ be any K-partition of the set of individual indices $\{1, 2, ..., N\}$. Let $\mathcal{G}_K$ denote the collection of such partitions. Let $\hat{\sigma}^2_{G^{(K)}} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in G_{K,k}} \sum_{t=1}^{T} [\tilde{Y}_{it} -$

$\tilde{Z}_{it}'\text{vec}\big(\hat{\omega}_{G_{K,k}}\big)]^2$, where $\text{vec}\big(\hat{\omega}_{G_{K,k}}\big) = \Big(\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\Big)^{+}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Y}_{it}$. The following assumptions are useful in the asymptotic development.

Assumption A4. *As* $(N,T)\to\infty$, $\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}} \xrightarrow{P} \underline{\sigma}^2 > \sigma_0^2$, *where* $\sigma_0^2 = \text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2$.

Assumption A5. *As* $(N,T)\to\infty$, $\rho_{NT}J\to 0$ *and* $\rho_{NT}NT\to\infty$.

Assumption A4 requires that all under-fitted models yield asymptotic mean square errors that are larger than σ_0^2, which is delivered by the true model. A5 imposes usual conditions for the consistency of model selection, namely, the penalty coefficient ρ_{NT} cannot shrink to zero either too fast or too slowly.

The following theorem justifies the use of (4.9) as a criterion to select K.

Theorem 4.5 *Suppose that Assumptions A1-A5 hold. Then* $P(\hat{K}=K_0)\to 1$ *as* $(N,T)\to\infty$.

Theorem 4.5 implies that the IC could determine the number of groups w.p.a.1. Of course, to implement it, one still needs to choose the tuning parameter ρ_{NT}. Motivated by BIC, we will set $\rho_{NT}=J_0\ln(NT)/(NT)$ in our simulations and application.

5 Extensions

In this section, we discuss several possible extensions of the time-varying panel structure model studied above.

5.1 Mixed Time-Varying Panel Structure Models

Consider the time-varying panel data models where some of the functional coefficients in β_{it}'s are common across all individuals whereas the others are group-specific. Write $\beta_{it} = (\beta_{it}^{(1)\prime},\beta_t^{(2)\prime})'$ where $\beta_{it}^{(1)}=\beta_i^{(1)}(t/T)$ is a $p_1\times 1$ vector of heterogenous functional coefficients that exhibits the following latent group structure

$$\beta_{it}^{(1)} = \sum_{k=1}^{K}\alpha_k(t/T)\cdot\mathbf{1}\{i\in G_k\},$$

and $\beta_t^{(2)}$ is $(p-p_1)\times 1$ vector of homogenous functional coefficients. Partition X_{it} conformably as $X_{it}=(X_{it}^{(1)\prime},X_{it}^{(2)\prime})'$. The model becomes

$$Y_{it}=\gamma_i+\beta_{it}^{(1)'}X_{it}^{(1)}+\beta_t^{(2)'}X_{it}^{(2)}+u_{it}, \tag{5.1}$$

where u_{it} and γ_i are defined as before. Our PSE method can be extended to this model straightforwardly. Given the spline basis system $B(v)$, we can approximate $\beta_i^{(1)}(v)$, $\beta^{(2)}(v)$, and $\alpha_k(v)$ by $\pi_i^{(1)'}B(v)$, $\pi^{(2)'}B(v)$, and $\omega_k'B(v)$, respectively. Let $\boldsymbol{\pi}^{(1)}=(\text{vec}(\pi_1^{(1)})',...,\text{vec}(\pi_N^{(1)})')'$, $\boldsymbol{\pi}^{(2)}=\text{vec}(\pi^{(2)})$, and $\boldsymbol{\omega}=(\text{vec}(\omega_1)',\ldots,\text{vec}(\omega_K)')'$. Now we can estimate $\boldsymbol{\pi}^{(1)}$, $\boldsymbol{\pi}^{(2)}$ and $\boldsymbol{\omega}$ by minimizing the following objective function:

$$Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^{(1)},\boldsymbol{\pi}^{(2)},\boldsymbol{\omega}\right)=Q_{1,NT}(\boldsymbol{\pi}^{(1)},\boldsymbol{\pi}^{(2)})+\frac{\lambda}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}(\pi_i^{(1)}-\omega_k)\right\| \tag{5.2}$$

where

$$Q_{1,NT}(\boldsymbol{\pi}^{(1)},\boldsymbol{\pi}^{(2)})=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{Y}_{it}-\tilde{Z}_{it}^{(1)\prime}\text{vec}(\pi_i^{(1)})-\tilde{Z}_{it}^{(2)\prime}\text{vec}(\pi^{(2)})\right\}^2,$$

$\tilde{Y}_{it}=Y_{it}-\frac{1}{T}\sum_{t=1}^{T}Y_{it}$, $\tilde{Z}_{it}^{(\ell)}=Z_{it}^{(\ell)}-\frac{1}{T}\sum_{t=1}^{T}Z_{it}^{(\ell)}$, $Z_{it}^{(\ell)}=X_{it}^{(\ell)}\otimes B(t/T)$ for $\ell=1,2$, $\tilde{V}_i=\{\text{diag}(\frac{J}{T}\tilde{Z}_i'\tilde{Z}_i)\}^{1/2}$, $\tilde{Z}_i=(\tilde{Z}_{i1},....,\tilde{Z}_{iT})'$, $\tilde{Z}_{it}=(\tilde{Z}_{it}^{(1)\prime},\tilde{Z}_{it}^{(2)\prime})'$, and $\tilde{\sigma}_i=\{\frac{1}{T}\sum_{t=1}^{T}[\tilde{Y}_{it}-\tilde{Z}_{it}'\text{vec}(\tilde{\pi}_i)]^2\}^{1/2}$, and $\tilde{\pi}_i=(\tilde{\pi}_i^{(1)},\tilde{\pi}^{(2)})$ is a preliminary estimate of $\pi_i=(\pi_i^{(1)},\pi^{(2)})$ obtained as in Section 3.2.

Let $\hat{\boldsymbol{\pi}}^{(1)}$=(vec$(\hat{\pi}_1^{(1)})'$,..., vec$(\hat{\pi}_N^{(1)})')'$, $\hat{\boldsymbol{\pi}}^{(2)}$=vec$(\hat{\pi}^{(2)})$, and $\hat{\boldsymbol{\omega}}=(\text{vec}(\hat{\omega}_1)',...,\ \text{vec}(\hat{\omega}_K)')'$ of $\boldsymbol{\pi}^{(1)}$, $\boldsymbol{\pi}^{(2)}$, and $\boldsymbol{\omega}$, respectively. We obtain the estimators of $\beta_i^{(1)}(v)$'s, $\beta^{(2)}(v)$, and $\alpha_k(v)$'s as follows:

$$\hat{\beta}_i^{(1)}(v)=\hat{\pi}_i^{(1)\prime}B(v),\ \hat{\beta}^{(2)}(v)=\hat{\pi}^{(2)\prime}B(v),\ \text{and}\ \hat{\alpha}_k(v)=\hat{\omega}_k'B(v), \tag{5.3}$$

where $i=1,...,N$, and $k=1,...,K$. Following the analyses in Sections 4.1-4.3, we can establish the asymptotic properties of the above estimators. In particular, we can establish the uniform consistency of the classification based on the PSE method and the oracle properties of $\hat{\alpha}_k(v)$ and $\hat{\beta}^{(2)}(v)$ and their post-Lasso versions. We omit the details due to the space constraint.

5.2 Unbalanced Panels

To broaden the applications of our model, we now consider an extension to unbalanced panels. Like Atak, Linton, and Xiao (2011) and for notational simplicity, we consider an unbalanced panel in which consecutive observations on individual units are available, but the number of time periods available may vary from unit to unit. The model becomes

$$Y_{it} = \gamma_i + \beta_{it}' X_{it} + u_{it}, \ i = 1, ..., N, \ t = t_i, ..., T_i, \tag{5.4}$$

where β_{it}'s have the latent group pattern in (2.2), and the other notations are defined as in Section 2. Let $\tau_i = T_i - t_i + 1$ and $n = \sum_{i=1}^{N} \tau_i$. Note that τ_i and n denote the total number of observations for individual i and the whole sample, respectively. Now we can estimate $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$ by minimizing the following objective function:

$$Q_{n,\lambda}^{(K)}(\boldsymbol{\pi},,\boldsymbol{\omega}) = Q_{1,NT}(\boldsymbol{\pi}) + \frac{\lambda}{N} \sum_{i=1}^{N} \tilde{\sigma}_i^{2-K} \prod_{k=1}^{K} \left\| \tilde{V}_i \text{vec}(\pi_i - \omega_k) \right\| \tag{5.5}$$

where

$$Q_{1,n}(\boldsymbol{\pi}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tau_i} \sum_{t=t_i}^{T_i} \left\{ \tilde{Y}_{it} - \tilde{Z}_{it}' \text{vec}(\pi_i) \right\}^2,$$

$\tilde{Y}_{it} = Y_{it} - \frac{1}{\tau_i} \sum_{t=t_i}^{T_i} Y_{it}$, $\tilde{Z}_{it} = Z_{it} - \frac{1}{\tau_i} \sum_{t=t_i}^{T_i} Z_{it}$, $Z_{it} = X_{it} \otimes B(t/T)$, $\tilde{V}_i = \{\text{diag}(\frac{J}{\tau_i} \tilde{Z}_i' \tilde{Z}_i)\}^{1/2}$, $\tilde{Z}_i = (\tilde{Z}_{it_i}, ..., \tilde{Z}_{iT_i})'$, $\tilde{\sigma}_i = \{\frac{1}{\tau_i} \sum_{t=t_i}^{T_i} [\tilde{Y}_{it} - \tilde{Z}_{it}' \text{vec}(\tilde{\pi}_i)]^2\}^{1/2}$, and $\tilde{\pi}_i$ is a preliminary estimate of π_i obtained as in Section 3.2. Let $\hat{\boldsymbol{\pi}}$ =(vec$(\hat{\pi}_1)'$, ..., vec$(\hat{\pi}_N)'$)$'$ and $\hat{\boldsymbol{\omega}}$ = (vec$(\hat{\omega}_1)'$, ..., vec$(\hat{\omega}_K)'$)$'$ of $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$, respectively. The formulae for the estimators $\hat{\beta}_i(v)$ and $\hat{\alpha}_k(v)$ of $\beta_i(v)$ and $\alpha_k(v)$ are the same as those given in (3.5). Define the estimated group $\hat{G}_k$ as in Section 4.1. The post-Lasso estimator of $\alpha_k(v)$ becomes

$$\hat{\alpha}_{\hat{G}_k}(v) = \hat{\omega}_{\hat{G}_k}' B(v), \tag{5.6}$$

where for $k = 1, ..., K$,

$$\text{vec}\left(\hat{\omega}_{\hat{G}_k}\right) = \left(\sum_{i \in \hat{G}_k} \frac{1}{\tau_i} \sum_{t=t_i}^{T_i} \tilde{Z}_{it} \tilde{Z}_{it}' \right)^+ \sum_{i \in \hat{G}_k} \frac{1}{\tau_i} \sum_{t=t_i}^{T_i} \tilde{Z}_{it} \tilde{Y}_{it}. \tag{5.7}$$

Let $\underline{T} \equiv \min_{1 \le i \le N} \tau_i$. To study the asymptotic properties of these estimators, we assume that $\underline{T} \to \infty$ and the conditions in Assumptions A1-A3 continue to hold with T replaced by $\underline{T}$. We need $\underline{T}\to\infty$ for the pointwise and mean square convergence results in Theorem 4.1, which are needed for the proofs of the uniform classification consistency and the oracle properties of $\hat{\alpha}_k(v)$ and its post-Lasso version. With some change in notation, the results in Theorem 4.1, Corollary 4.2, and Theorems 4.3-4.4 continue to hold. In particular, the results in Theorem 4.4 become:

(i) $\sqrt{N_k\underline{T}/J}\mathbb{S}_k^{-1/2}\left[\hat{\alpha}_k\left(v\right)-\alpha_k\left(v\right)\right]\xrightarrow{D}N\left(0,\mathbb{I}_p\right),$

(ii) $\sqrt{N_k\underline{T}/J}\mathbb{S}_k^{-1/2}\left[\hat{\alpha}_{\hat{G}_k}\left(v\right)-\alpha_k\left(v\right)\right]\xrightarrow{D}N\left(0,\mathbb{I}_p\right),$

where $\mathbb{S}_k=\left(\mathbb{I}_p\otimes B\left(v\right)\right)'\left(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\{\frac{J}{N_k}\sum_{i\in G_k^0}\frac{T}{\tau_i^2}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\left(\mathbb{I}_p\otimes B\left(v\right)\right),$ and $\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}=\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{\tau_i}\sum_{t=t_i}^{T_i}\tilde{Z}_{it}\tilde{Z}_{it}'$ for $k=1,...,K.$

5.3 Panels with Cross-Sectional Dependence

We can also allow for cross-sectional dependence in our model. A popular way to introduce the cross-sectional dependence is via the use of the interactive fixed effects:

$$Y_{it}=\beta_{it}'X_{it}+\gamma_i'F_t+u_{it}, \tag{5.8}$$

where γ_i and F_t denote an $R\times 1$ vector of factor loadings and common factors, respectively, both of which can be correlated with $\{X_{it}\}$, β_{it}'s have the latent group pattern in (2.2), and the other notations are defined as in Sections 2. When $R=1$, $F_t=1$, the model in (5.8) becomes the model in (2.1) with additive fixed effects. Let $F=(F_1,...,F_T)'$ and $\Lambda=(\gamma_1,...,\gamma_N)'$. Following Bai and Ng (2002), Moon and Weidner (2015), and Su and Ju (2017), we impose the identification restrictions: $T^{-1}F'F/T=\mathbb{I}_R$, $\Gamma'\Gamma$ =diagonal with descending diagonal elements. Let $Y_i\equiv(Y_{i1},\ldots,Y_{iT})'$ and $Z_i\equiv(Z_{i1},\ldots,Z_{iT})'$ where recall that $Z_{it}=X_{it}\otimes B\left(t/T\right)$. We propose to estimate $\{\pi_i\}$, $\{\omega_k\}$, F, and Γ by minimizing the following penalized objective function

$$Q_{0NT,\lambda}^{(K)}\left(\boldsymbol{\pi},\boldsymbol{\omega},F,\Gamma\right)=Q_{0,NT}\left(\boldsymbol{\pi},F,\Gamma\right)+\frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\mathrm{vec}(\pi_i-\omega_k)\| \tag{5.9}$$

where $Q_{0,NT}\left(\boldsymbol{\pi},F,\Gamma\right)=\frac{1}{NT}\sum_{i=1}^{N}\|Y_i-Z_i\mathrm{vec}(\pi_i)-F\gamma_i\|^2$, and F and Γ satisfy the above identification restrictions. Following Moon and Weidner (2015) and Su and Ju (2017), we can concentrate Γ and F out in turn and obtain the profile objective function

$$Q_{NT,\lambda}^{(K)}(\boldsymbol{\pi},\boldsymbol{\omega}) = Q_{1,NT}(\boldsymbol{\pi}) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\text{vec}(\pi_i - \omega_k)\|, \tag{5.10}$$

where $Q_{1,NT}(\boldsymbol{\pi}) = \frac{1}{T}\sum_{r=R+1}^{T}\mu_r\left[\frac{1}{N}\sum_{i=1}^{N}(Y_i - Z_i\text{vec}(\pi_i))(Y_i - Z_i\text{vec}(\pi_i))'\right]$ and $\mu_r(A)$ denotes the rth largest eigenvalue of A by counting multiple eigenvalues multiple times.

Minimizing the criterion function in (5.10) produces the C-Lasso estimators $\hat{\boldsymbol{\pi}}$ =(vec$(\hat{\pi}_1)'$, ...,vec$(\hat{\pi}_N)')'$ and $\hat{\boldsymbol{\omega}}$ = (vec$(\hat{\omega}_1)'$, ..., vec$(\hat{\omega}_K)')'$ of $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$. The estimators $\hat{F}$ and $\hat{\Gamma}$ of F and Γ are obtained as the solutions to the following eigenvalue problem:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - Z_i\text{vec}(\hat{\pi}_i))(Y_i - Z_i\text{vec}(\hat{\pi}_i))'\right]\hat{F} = \hat{F}V_{NT} \text{ and } \hat{\Gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_N)', \tag{5.11}$$

where V_{NT} is a diagonal matrix consisting of the R largest eigenvalues of the above matrix in the square bracket, arranged in descending order, and $\hat{\gamma}_i = T^{-1}\hat{F}'(Y_i - Z_i\text{vec}(\hat{\pi}_i))$. The formulae for the estimators $\hat{\beta}_i(v)$ and $\hat{\alpha}_k(v)$ of $\beta_i(v)$ and $\alpha_k(v)$ are the same as those given in (3.5). Following the technical analyses in Su and Ju (2017) and those in Sections 4.1-4.3, we can establish the asymptotic properties of the above estimators. In particular, we can establish the uniform consistency of the classification and the oracle properties of $\hat{\alpha}_k(v)$ and its post-Lasso version. For brevity, we omit the details.

6 Monte Carlo Study and Empirical Illustration

6.1 Monte Carlo Study

In this section, we evaluate the finite sample performance of the information criterion in determining the number of groups and the C-Lasso and post-Lasso estimates.

6.1.1 Data Generating Processes

We consider three data generation processes (DGPs). In all DGPs, the fixed effect γ_i and the idiosyncratic error u_{it} follow the standard normal distribution and are mutually independent across both i and t. The observations in each DGP are drawn from three groups with

$N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$. We consider four combinations of the sample sizes with $N = 50, 100$ and $T = 40, 80$.

DGP 1 (Trending panel structure model) Y_{it} is generated via $Y_{it} = \gamma_i + \beta_i^0\,(t/T) + u_{it}$, where

$$\beta_i^0\,(v) = \begin{cases} \alpha_1^0(v) = 6F(v, 0.5, 0.1) & \text{if } i \in G_1^0 \\ \alpha_2^0(v) = 6\left[2v - 6v^2 + 4v^3 + F(v; 0.7, 0.05)\right] & \text{if } i \in G_2^0 \\ \alpha_3^0(v) = 6\left[4v - 8v^2 + 4v^3 + F(v; 0.6, 0.05)\right] & \text{if } i \in G_3^0 \end{cases}, \tag{6.1}$$

$F(\cdot; \mu, s) = \{1 + \exp[-(\cdot - \mu)/s]\}^{-1}$ denotes the cumulative distribution function of the logistic distribution with location and scale parameters given by μ and s, respectively.

DGP 2 (Time-varying panel structure model with an exogenous regressor) Y_{it} is generated via $Y_{it} = \gamma_i + \beta_{i,1}^0\,(t/T) + \beta_{i,2}^0\,(t/T)\,X_{it} + u_{it}$, where $\{X_{it}\}$ is an IID $N(0, 1)$ sequence, $\beta_{i,1}^0\,(v) = \frac{1}{2}\beta_i^0\,(v)$ with $\beta_i^0\,(v)$ given in (6.1),

$$\beta_{i,2}^0\,(v) = \begin{cases} \alpha_{1,2}^0(v) = 3\left[2v - 4v^2 + 2v^3 + F(v; 0.6, 0.1)\right] & \text{if } i \in G_1^0 \\ \alpha_{2,2}^0(v) = 3\left[v - 3v^2 + 2v^3 + F(v; 0.7, 0.04)\right] & \text{if } i \in G_2^0 \\ \alpha_{3,2}^0(v) = 3\left[0.5v - 0.5v^2 + F(v; 0.4, 0.07)\right] & \text{if } i \in G_3^0 \end{cases}, \tag{6.2}$$

and F is defined as above. Here, the first element in the group-specific parameter vector $\alpha_k^0\,(\cdot)$ is given by $\alpha_{k,1}^0\,(\cdot) = \frac{1}{2}\alpha_k^0\,(\cdot)$ with $\alpha_k^0\,(\cdot)$ defined in (6.1). The left and right panels of Figure 1 depict the group-specific time trends $\alpha_{k,1}^0(\cdot)$ and $\alpha_{k,2}^0(\cdot)$ for $k = 1, 2, 3$, respectively.

DGP 3 (Time-varying dynamic panel structure model) Y_{it} is generated via $Y_{it} = \gamma_i + \beta_{i,3}^0\,(t/T)\,Y_{it-1} + u_{it}$, where

$$\beta_{i,3}^0\,(v) = \begin{cases} \alpha_{1,3}^0(v) = \frac{3}{2}\left[-0.5 + 2v - 5v^2 + 2v^3 + F(v; 0.6, 0.03)\right] & \text{if } i \in G_1^0 \\ \alpha_{2,3}^0(v) = \frac{3}{2}\left[-0.5 + v - 3v^2 + 2v^3 + F(v; 0.2, 0.04)\right] & \text{if } i \in G_2^0 \\ \alpha_{3,3}^0(v) = \frac{3}{2}\left[-0.5 + 0.5v - 0.5v^2 + F(v; 0.8, 0.07)\right] & \text{if } i \in G_3^0 \end{cases}, \tag{6.3}$$

and F is defined as above.

In addition, we also check the performance of our method when the error terms exhibit weak cross-sectional dependence and when the number of groups is large. The simulation results are quite similar to those reported below. Due to the space limit, we do not reports these results in this paper.

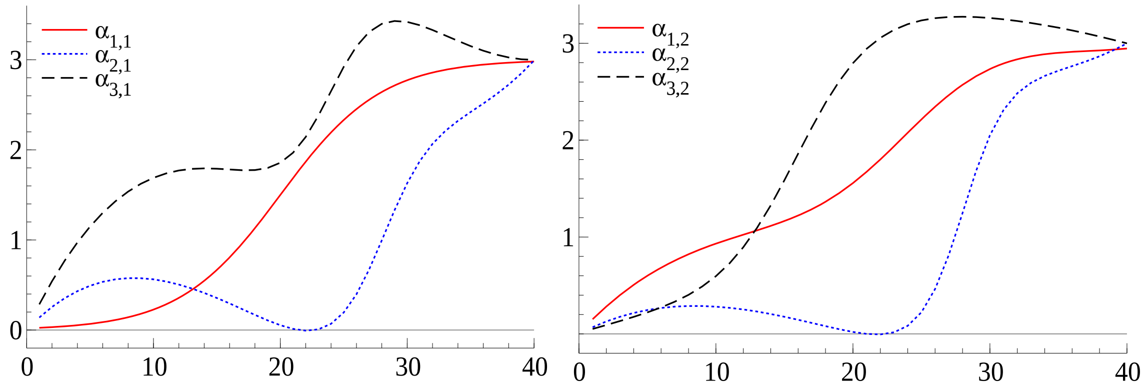

Figure 1: The plots of the group-specific functional coefficients in DGP 2 (Left panel: solid, dotted and dashed lines for $\alpha^0_{1,1}(\cdot)$, $\alpha^0_{2,1}(\cdot)$, and $\alpha^0_{3,1}(\cdot)$, respectively in DGP 2; Right panel: solid, dotted and dashed lines for $\alpha^0_{1,2}(\cdot)$, $\alpha^0_{2,2}(\cdot)$, and $\alpha^0_{3,2}(\cdot)$, respectively in DGP 2)

6.1.2 Determination of the Number of Groups

In this subsection, we assess the performance of (4.9) in determining the number of groups. We set $J_0 = \lfloor (NT)^{1/6} \rfloor$, the number of knots in the cubic B-spline approximation, where $\lfloor A \rfloor$ denotes the integer part of A. We set $\lambda = c_\lambda (NT)^{-(2K+3)/24}$ and consider various values of c_λ to examine the sensitivity of the IC's performance to the choice of λ. We consider $c_\lambda = 1, 2, 4$ but only report the results for $c_\lambda = 1$ here to save space. The results for $c_\lambda = 2$ and 4 are quite similar and available upon request from the authors.

For each DGP, we simulate 200 data sets for each of the four combinations of N and T. We evaluate the IC for $K = 1, 2, ..., K_{\max}$ with $K_{\max} = 6$ and select the optimal number of groups by minimizing the IC in (4.9). Table 1 reports the empirical probability that a specific number of groups is selected based on 200 replications. As shown in the table, our IC works fairly well.

6.1.3 Classification and Estimation

As shown in the previous subsection, the IC in Section 4.4 works fairly well in finite samples. In this subsection, we assume that the number of groups is known and focus on the classification and estimation.

We set the tuning parameter λ as above. We set the initial values of π_i's to be $\tilde{\pi}_i$'s

Table 1: The performance of information criterion in determining the number of groups

DGP	N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
1	50	40	0.000	0.015	0.985	0.000	0.000	0.000
	50	80	0.000	0.000	0.985	0.015	0.000	0.000
	100	40	0.000	0.000	0.995	0.005	0.000	0.000
	100	80	0.000	0.000	1.000	0.000	0.000	0.000
2	50	40	0.000	0.075	0.925	0.000	0.000	0.000
	50	80	0.000	0.000	0.990	0.010	0.000	0.000
	100	40	0.000	0.010	0.990	0.000	0.000	0.000
	100	80	0.000	0.000	0.980	0.020	0.000	0.000
3	50	40	0.000	0.120	0.880	0.000	0.000	0.000
	50	80	0.000	0.030	0.970	0.000	0.000	0.000
	100	40	0.000	0.020	0.980	0.000	0.000	0.000
	100	80	0.000	0.005	0.995	0.000	0.000	0.000

Note: The main entries are the empirical probability that a specific number of groups is selected based on 200 replications.

and those of ω_k's to be zero. We have also tried other initial values and found that the classification and estimation results are quite similar to those reported here, suggesting the robustness of our algorithm to the initial values of parameters.

We run 200 replications for each DGP and classify individual i into group k if $\|\hat{\beta}_i - \hat{\alpha}_k\|$ achieves the minimum. To measure the accuracy of classification, we report two types of classification errors as defined in Section 4.2, i.e., $\bar{P}(\hat{E}) = \frac{1}{N}\sum_{i=1}^{N} \hat{P}(\cup_{k=1}^{K} \hat{E}_{k,NT,i})$ and $\bar{P}(\hat{F}) = \frac{1}{N}\sum_{i=1}^{N} \hat{P}(\cup_{k=1}^{K} \hat{F}_{k,NT,i})$, where $\hat{P}$ denotes the empirical average probabilities across 200 replications. Table 2 reports the classification errors. The results with different c_λ's are quite similar, indicating the robustness of our algorithm to the choice of tuning parameter. Moreover, the classification errors $\bar{P}(\hat{E})$ and $\bar{P}(\hat{F})$ are all below 3% for each scenario of the first two DGPs. The classification errors are a little bit large for the dynamic panel data models, but are still acceptable. In particular, all of them shrink toward zero quickly as T increases.

For the estimation, Figure 2 depicts the three true group-specific trends and their post-Lasso estimates in DGP 2 for the case $N = 100, T = 40$ based on 200 replications. As shown in Figure 2, the fitted trends approximate the true trends pretty well, indicating the excellent behavior of our estimation procedure. To measure the accuracy of estimation for the group-specific functional coefficients, we define the weighted root-mean-square-error

Table 2: Two types of classification error in percentages

DGP	N	T	$c_\lambda = 1$		$c_\lambda = 2$		$c_\lambda = 4$	
			$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$
1	50	40	0.460	0.434	0.380	0.357	0.600	0.570
	50	80	0.110	0.080	0.000	0.000	0.010	0.009
	100	40	0.580	0.606	0.855	0.747	0.425	0.415
	100	80	0.015	0.015	0.005	0.005	0.010	0.010
2	50	40	2.930	2.734	2.870	2.617	2.100	1.959
	50	80	0.570	0.389	0.340	0.289	0.230	0.114
	100	40	1.645	1.547	2.665	2.460	1.765	1.700
	100	80	0.800	0.984	0.120	0.114	0.070	0.068
3	50	40	6.730	6.542	6.535	6.178	7.720	6.917
	50	80	2.175	2.052	1.995	1.873	2.230	2.208
	100	40	5.585	5.692	5.360	5.294	5.970	5.973
	100	80	1.135	1.087	1.050	1.008	1.190	1.084

Table 3: Root mean squared errors of the C-Lasso and post-Lasso estimates

DGP	coeff	N	T	oracle	$c_\lambda = 1$		$c_\lambda = 2$		$c_\lambda = 4$	
					C-Lasso	post-Lasso	C-Lasso	post-Lasso	C-Lasso	post-Lasso
1	α_k^0	50	40	0.166	0.218	0.167	0.233	0.167	0.205	0.169
		50	80	0.150	0.165	0.151	0.170	0.150	0.165	0.150
		100	40	0.153	0.220	0.155	0.261	0.156	0.206	0.154
		100	80	0.116	0.141	0.116	0.214	0.116	0.160	0.116
2	$\alpha_{k,1}^0$	50	40	0.117	0.143	0.123	0.175	0.126	0.141	0.119
		50	80	0.096	0.111	0.099	0.138	0.096	0.110	0.098
		100	40	0.097	0.126	0.099	0.185	0.101	0.133	0.100
		100	80	0.076	0.106	0.082	0.180	0.076	0.110	0.076
	$\alpha_{k,2}^0$	50	40	0.120	0.144	0.126	0.165	0.128	0.141	0.123
		50	80	0.096	0.112	0.101	0.127	0.096	0.110	0.099
		100	40	0.097	0.122	0.099	0.185	0.102	0.127	0.099
		100	80	0.065	0.096	0.072	0.183	0.065	0.099	0.065
3	$\alpha_{k,3}^0$	50	40	0.297	0.418	0.397	0.472	0.453	0.432	0.403
		50	80	0.201	0.323	0.292	0.334	0.304	0.318	0.301
		100	40	0.210	0.394	0.383	0.397	0.375	0.408	0.399
		100	80	0.146	0.307	0.271	0.311	0.283	0.302	0.284

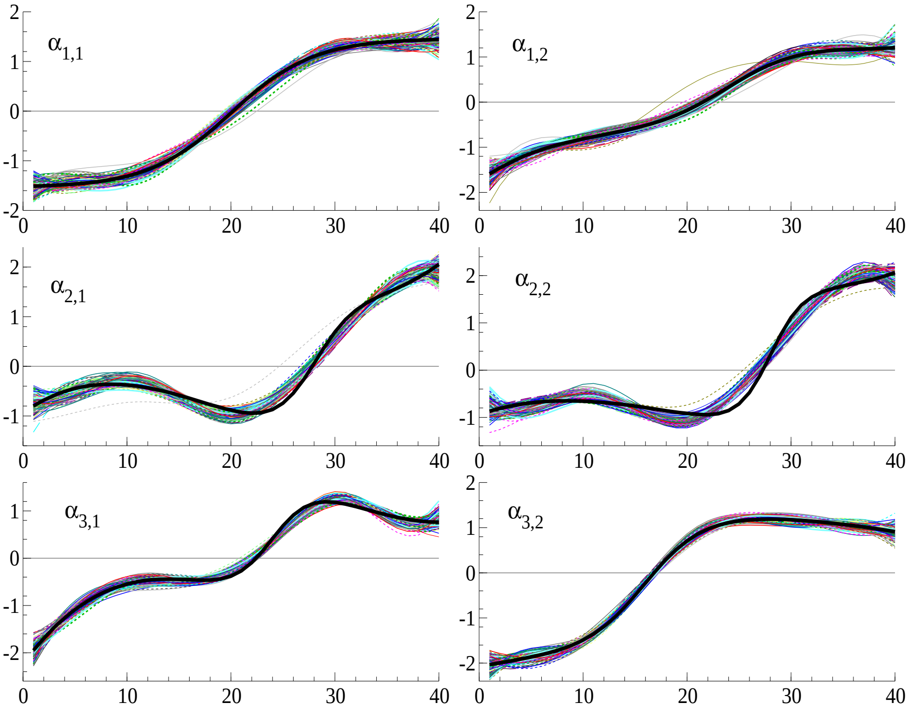

Figure 2: True trends (the heavy black line) and the post-Lasso estimators for DGP2 ($N = 100, T = 40$)

(RMSE) of the estimates $\hat{\alpha}_k(t/T)$'s in DGP 1 for each replication as follows

$$RMSE(\hat{\alpha}_{\cdot}) = \frac{1}{N}\sum_{k=1}^{3} N_k RMSE(\hat{\alpha}_k),$$

where $RMSE(\hat{\alpha}_k) = \{\frac{1}{T}\sum_{t=1}^{T}[\hat{\alpha}_k(t/T) - \alpha_k^0(t/T)]^2\}^{1/2}$ for $k = 1, 2, 3$. The weighted RMSEs of the estimates of $\alpha_{k,1}^0(t/T)$ and $\alpha_{k,2}^0(t/T)$ in DGP 2 are similarly defined. Table 3 reports the average of these RMSEs across 200 replications for both the C-Lasso and post-Lasso estimators for $c_\lambda = 1, 2, 4$, in comparison with the oracle estimators. As shown in Table 3, the RMSEs are quite similar for different choices of c_λ and generally decline as T increases for fixed N. The RMSEs of the post-Lasso estimators are less than those of the C-Lasso estimators in all cases and they are close to those of the oracle estimators when $T = 80$. This suggests that the post-Lasso estimators tend to outperform the C-Lasso estimators and would be recommended for practical use.

6.2 Empirical Illustration

As a key indicator of a country's standard of living, GDP per capita has been one of the most important variables in economics; see, e.g., Solow (1956), Cass (1965), and Barro (1991, 1996). It not only provides a useful statistic for comparison of wealth across countries but also describes the development of a particular country. However, the exact realization of GDP per capita is not very useful in comparison due to the existence of the short term fluctuations. In fact, policy makers often target on long-lasting changes rather than short transitory fluctuations. This prompts us to extract the trend of GDP per capita, which can capture the medium-to long-term changes and have some implications on economic modeling, testing and forecasting. For example, most of the existing literature assumes a linear trend behavior for the GDP per capita when testing trend stationarity against unit root; see, e.g., Fleissig and Strauss (1999) and Lluís Carrion-I-Silvestre, Barrio-Castro, and López-Bazo (2005). If the underlying trend is nonlinear in fact, then the conclusions can be misleading.

In this section, we use our time-varying panel data model with latent structure to estimate the heterogeneous trending behavior of GDP per capita across countries. In comparison with Robinson's (2012) nonparametric panel trend model, our model allows for unobserved cross-sectional heterogeneity. This is important as it is hard to believe the GDP per capita for all countries exhibit the same trend over time. Although macroeconomists have some consensus that globalization leads to the synchronization of business cycles across countries (Kose, Prasad, and Terrones 2003), it is unrealistic to assume all the countries share the same trend. In fact, a stream of empirical studies confirm the cross-country divergence rather than convergence implied by the neoclassical growth models (Barro, 1991), and thus provide ample evidences on the cross-sectional heterogeneity. To account for the cross-sectional heterogeneity, applied researchers usually select a small group of countries (e.g., OECD countries) that they think would share slope homogeneity, and then conduct statistical analysis for the selected countries. However, such a selection appears arbitrary which may further induce misleading results. As mentioned above, our model provides a data-driven classification before we embark on the estimation and inference procedure for the group-specific trending behavior. Hence, it is useful to extract the group-specific trends of GDP per capita across countries based on our new methodology.

6.2.1 Data and Setting

Denote the GDP per capita as Y_{it}. Then we estimate the following trending panel structure model

$$\log Y_{it} = \gamma_i + \beta_i(t/T) + u_{it}$$

using the annual data from 1960 to 2012 for as many countries as possible. We obtain the GDP per capita data from Federal Reserve Economic Data (FRED), measured in terms of 2005 U.S. dollars. By deleting countries with missing observations, we obtain a balanced panel that contains $N = 91$ countries and $T = 53$ observations for each country. As we are interested in the common trend across countries, we take logarithm for the data. By taking logarithm, the slope of the trend could be roughly interpreted as the growth rate of GDP per capita up to a scaling factor T. The data series are depicted in Figure 3. To show the path of the data more clearly, we report the demeaning data. It is obvious that the time path of GDP per capita exhibits noticeable heterogeneity. The whole world realization, which is marked by the thick black curve in the figure, is not a reasonable representative of the economic development.

We estimate the trending panel structure model in (2.5) by using the iterative algorithm

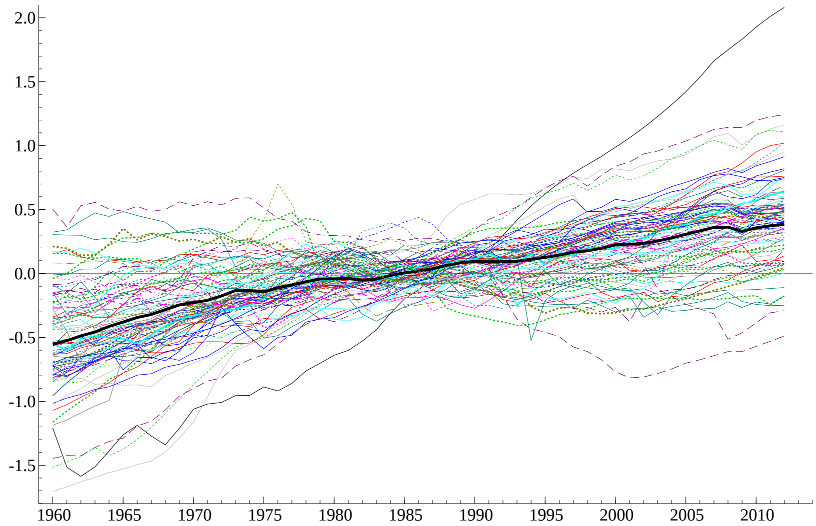

Figure 3: GDP per capita (logarithm and demean) for 91 countries between 1960 and 2012 (logarithm and demeaned, the thick black curve denotes the value for the whole world)

introduced in the online Appendix C. To implement the penalized least squares estimation with cubic B-spline approximation, we set the number of knots (J_0) and tuning parameter λ as in the simulation section.

6.2.2 Estimation Results

To determine the appropriate number of groups, we choose K to minimize the information criterion in Section 4.4. Table 4 reports the ICs for the number of group $K = 1, 2, ..., 6$ with different tuning parameters $c_\lambda = 0.5, 1, 2, 4$. The results show that the IC is robust to the tuning parameter and always achieves the minimum when $K = 4$. Figure 4 depicts the estimated trends for the four estimated groups and Figure 5 reports the realization of GDP per capita (logarithm and demeaned) and the trend for each group. To save space, we do not report the detailed estimation results here. A detailed report for the empirical results and discussions can be found in the online Appendix D.

Table 4: The information criterion for different numbers of groups

$c_\lambda \setminus K$	1	2	3	4	5	6
0.5	-2.584	-3.174	-3.344	**-3.605**	-3.234	-2.500
1	-2.584	-3.174	-3.360	**-3.559**	-2.922	-2.504
2	-2.584	-3.174	-3.375	**-3.407**	-3.237	-2.555
4	-2.584	-3.174	-3.344	**-3.697**	-3.409	-2.687

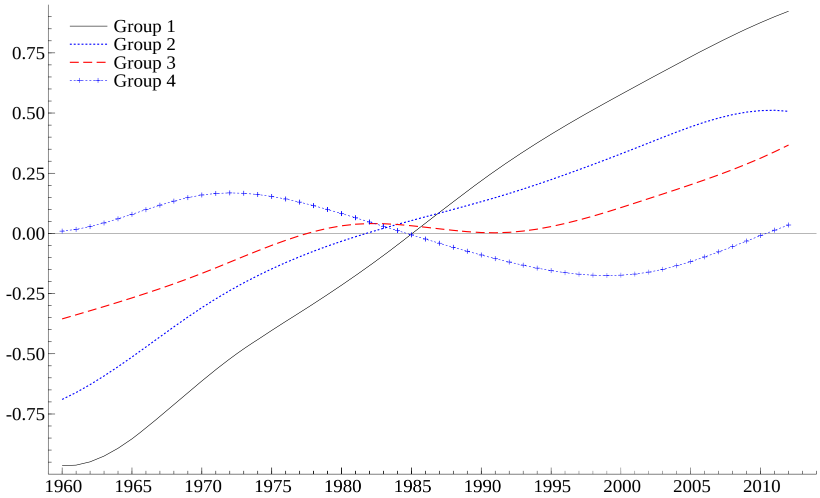

Figure 4: The estimated trends for the four estimated groups of economies

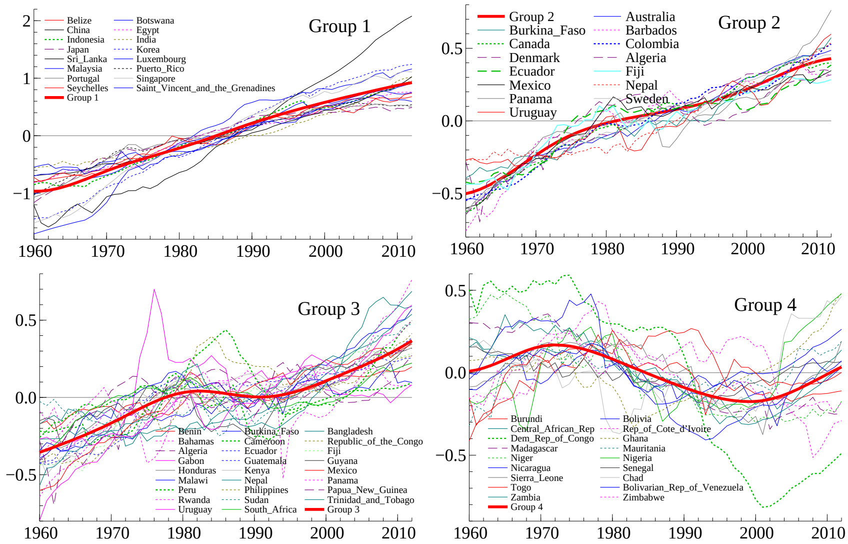

Figure 5: The GDP per capita (logarithm and demean) for countries in each group and the estimated group-specific trend (thick solid curve)

6.3 Further Discussion on Potential Applications

The proposed time-varying panel data model with latent structures could capture the smooth structural changes over time and the individuals' heterogeneity across groups simultaneously. The model is flexible and hence is expected to have much broad applications in empirical study. As mentioned before, changes induced by policy switch, preference change, and technology progress can cause structural changes of the functional relationships between economic variables. Besides, individual units sampled from different backgrounds are expected to have heterogeneous features. To handle the individual heterogeneity, many empirical studies classify units to different groups based on some external criterion. For example, in macroeconomic studies involving countries, researchers often consider the OECD countries and the emerging economies separately. In microeconomic studies, individuals are usually classified into low income group, middle income group and high income group. By adopting our method, one does not need to classify units into different groups *a priori*. Our method could identify individuals' membership endogenously. Here, we discuss two potential applications.

The first example is the energy intensity. Energy intensity is a measure of the energy efficiency of a nation's economy. It is calculated as units of energy per unit of GDP. High energy intensities indicate a high price or cost of converting energy into GDP. The trend of the energy intensity reveals the changes of the economic energy efficiency. Due to the different stages of economic development that different countries attain, the trend of energy intensity varies across countries. Hence, we can consider the following trending panel structure model to estimate the trend of energy intensity for various countries:

$$y_{it} = \gamma_i + \beta_i(t/T) + u_{it}$$

where γ_i is the individual effect and $\beta_i(t/T)$ satisfies the latent group structure in our paper.

The second example is the beneficial effects of foreign direct investment (FDI) on economic growth in host countries over a long period of time. As mentioned before, the relationships between variables tend to change during a long period. In addition, due to the difference of absorptive capacities in different host countries, the FDI effects tend to be heterogeneous. To capture the time-varying relationship and the cross-country heterogeneous absorptive capacities simultaneously, we consider the following model:

$$y_{it} = \alpha_i + \beta_{it}^{(1)}(FDI/Y)_{it} + \beta_t^{(2)} \log(DI/Y)_{it} + \beta_t^{(3)} n_{it} + \beta_t^{(4)} h_{it} + \beta_t^{(5)} \left((FDI/Y)_{it} \times h_{it}\right) + u_{it}$$

where y_{it} denotes the growth rate of GDP per capita in country/region i during the period t, n_{it} is the logarithm of population growth rate, h_{it} is the human capital, and α_i is the individual effect used to control the unobserved country-specific heterogeneity. Here, FDI and DI refer to foreign direct investment and domestic investment, respectively; Y represents the total output. Hence, $(FDI/Y)_{it}$ denotes the average ratio between the FDI and the total output during the period t in country/region i, and $(DI/Y)_{it}$ is defined in the same fashion for the domestic investment. In the model, $\beta_{it}^{(1)}$ exhibits the latent group structure in (2.2), and $\beta_t^{(j)} = \beta^{(j)}(t/T)$, $j = 2, 3, ..., 5$, are homogeneous functional coefficients. This model is the mixed time-varying panel structure model that can be estimated by using the technique given in Section 5.1. Alternatively, we can also allow $\beta_t^{(j)}, j = 2, 3, ..., 5$, to be heterogeneous and have the latent group structure. In either case, the model extends the typical empirical growth equation

$$y_{it} = \alpha_i + \beta^{(1)}(FDI/Y)_{it} + \beta^{(2)} \log(DI/Y)_{it} + \beta^{(3)} n_{it} + \beta^{(4)} h_{it} + \beta^{(5)} \left((FDI/Y)_{it} \times h_{it}\right) + u_{it}.$$

See Kottaridi and Stengos (2010) and Cai, Chen and Fang (2014) and the references therein.

7 Conclusion

In this paper we propose a time-varying panel data model with latent group structures to capture individual heterogeneity and smooth structural changes over time simultaneously. We focus on the penalized sieve estimation (PSE) of such a model where the penalty term is constructed to achieve simultaneous classification and estimation. The PSE achieves the uniform classification consistency and oracle property. We also propose a BIC-type information criterion to determine the unknown number of groups. Simulations are conducted to evaluate the finite sample performance of the proposed information criterion and PSE method. We apply our method to study the heterogeneous trending behavior of GDP per capita across 91 countries for the period 1960-2012 and find four latent groups.

Several extensions are possible. First, one can consider general functional coefficient panel data models with latent group structures where the coefficients are functions of certain random covariates. More generally, one can consider other types of nonparametric or semiparametric panel data models (e.g., the partially linear single-index panel data model of Chen, Gao, and Li 2013) with latent group structures. Second, as discussed in Section 5.3 we can also allow for cross-sectional dependence in our model. But the asymptotic theory is extremely involved and we leave it for future research.

Mathematical Appendix

A Proofs of the Results in Section 4

We first state some lemmas that are used in the proof of the main results in Section 4. The proofs of these lemmas are available in the online supplementary appendix.

We use various properties of B-splines in our proofs. Recall that $B(v) = (B_{-d+1}(v), B_{-d+2}(v), ..., B_{J_0}(v))'$ for $v \in [0,1]$ and $J = J_0 + d$. The B-splines $B = (B_j)$ of order d have the following properties:

(i) $B_j(v) \geq 0$ for each $j = -d+1, ..., J_0$ and $v \in [0,1]$, and $B_j(v)$ vanishes outside the interval $[v_j, v_{j+d})$; see de Boor (2001, p.91)

(ii) $\sum_{j=-d+1}^{J_0} B_j(v) = 1$ for $v \in [0,1]$; see de Boor (2001, p.96).

(iii) There are two positive constants M_1 and M_2 such that

$$\frac{M_1}{J} \|\mathbf{c}\|^2 \leq \int \{\mathbf{c}' B(v)\}^2 dv \leq \frac{M_2}{J} \|\mathbf{c}\|^2 \text{ for all } \mathbf{c} \in \mathbb{R}^J. \tag{A.1}$$

See DeVore and Lorentz (1993, p.145) or de Boor (2001, p.133).

Clearly, (i)-(ii) imply that $B_j(v)$ is uniformly bounded on $[0,1]$ and $\int_0^1 B_j(v) dv = O(J^{-1})$ uniformly in j. (iii) implies that the maximum and minimum eigenvalues of $J \int_0^1 B(v) B(v)' dv$ are bounded from above by M_2 and from below by M_1, respectively.

For $\mathbf{g}^{(1)}(t/T) = (g_1^{(1)}(t/T), ..., g_p^{(1)}(t/T))'$ and $\mathbf{g}^{(2)}(t/T) = (g_1^{(2)}(t/T), ..., g_p^{(2)}(t/T))'$, define the empirical and theoretical inner products respectively as $< \mathbf{g}^{(1)}, \mathbf{g}^{(2)} >_{i,T} = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{g}^{(1)}(t/T)' X_{it}) \left(X_{it}' \mathbf{g}^{(2)}(t/T)\right)$ and $< \mathbf{g}^{(1)}, \mathbf{g}^{(2)} >_i = E\left[< \mathbf{g}^{(1)}, \mathbf{g}^{(2)} >_{i,T}\right]$. Denote the corresponding norms as $\|\cdot\|_{i,T}$ and $\|\cdot\|_i$, respectively.

Lemma A.1 *Suppose that Assumption A1 holds. Let* $\mathbb{G} \equiv \{g(\cdot) = \omega' B(\cdot),\ \omega \in \mathbb{R}^J\}$ *and* $g_l(v) = \pi_{i,l}' B(v)$ *for* $l = 1, ..., p$. *Let* $\mathbf{g}(v) = (g_1(v), ..., g_p(v))'$. *Then* $\|\mathbf{g}\|_i^2 \asymp \sum_{l=1}^p \|g_l\|_2^2 \asymp J^{-1} \|vec(\pi_i)\|^2$.

Lemma A.2 *Suppose that Assumption A1 holds. Let* $\mathbb{G}^{\otimes p}$ *denote the collection of vectors of spline functions* $\mathbf{g} = (g_1, ..., g_p)'$ *with* $g_l \in \mathbb{G}$ *defined as above. Then for any* $\epsilon > 0$

(i) $P\left(\max_{1 \leq i \leq N} \sup_{\mathbf{g} \in \mathbb{G}^{\otimes p}} \left| \frac{\frac{1}{T}\sum_{t=1}^T \left[\mathbf{g}(t/T)' X_{it}\right]^j}{\frac{1}{T}\sum_{t=1}^T E\left[\mathbf{g}(t/T)' X_{it}\right]^j} - 1 \right| > \epsilon \right) = o(N^{-1})$ *for* $j = 1, 2$;

(ii) $P\left(\sup_{\mathbf{g} \in \mathbb{G}^{\otimes p}} \left| \frac{\sum_{i=1}^N \|\mathbf{g}\|_{i,T}^2}{\sum_{i=1}^N \|\mathbf{g}\|_i^2} - 1 \right| > \epsilon \right) = P\left(\sup_{\mathbf{g} \in \mathbb{G}^{\otimes p}} \left| \frac{\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{g}(t/T)' X_{it}\right]^2}{\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T E\left[\mathbf{g}(t/T)' X_{it}\right]^2} - 1 \right| > \epsilon \right)$ $= o(N^{-1})$.

Lemma A.3 *Suppose that Assumption A1 holds. Let* $\hat{Q}_{i,\tilde{z}\tilde{z}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$. *Recall* $\tilde{V}_i = [diag(J\hat{Q}_{i,\tilde{z}\tilde{z}})]^{1/2}$. *Then there exist two positive constants* $\underline{c}_{zz}$ *and* $\bar{c}_{zz}$ *that do not depend on* N, T*, or* J *such that (i)* $P(\underline{c}_{zz} \leq \min_{1\leq i\leq N}\mu_{\min}(J\hat{Q}_{i,\tilde{z}\tilde{z}}) \leq \max_{1\leq i\leq N}\mu_{\max}(J\hat{Q}_{i,\tilde{z}\tilde{z}}) \leq \bar{c}_{zz}) = 1-o(N^{-1})$*, and (ii)* $P(\underline{c}_{zz}^{1/2} \leq \min_{1\leq i\leq N}\mu_{\min}(\tilde{V}_i) \leq \max_{1\leq i\leq N}\mu_{\max}(\tilde{V}_i) \leq \bar{c}_{zz}^{1/2}) = 1-o(N^{-1})$.

Lemma A.4 *Suppose that Assumption A1 holds. Let* $a_{it} = \beta_i^0(t/T) - \pi_i^{0\prime}B(t/T)$ *and* $\eta_{it} = X_{it}'a_{it}$. *Let* $\hat{Q}_{i,\tilde{z}\tilde{\zeta}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{\zeta}_{it}$ *for* $\zeta = e, \eta$*, and* u*, where* $\tilde{e}_{it} = e_{it} - \frac{1}{T}\sum_{t=1}^{T}e_{it}$, $\tilde{\eta}_{it}$ *and* $\tilde{u}_{it}$ *are analogously defined. Then (i)* $||\hat{Q}_{i,\tilde{z}\tilde{e}}|| = O_P(J^{-\gamma-1/2} + T^{-1/2})$*, (ii)* $\frac{1}{N}\sum_{i=1}^{N}||\hat{Q}_{i,\tilde{z}\tilde{e}}||^2 = O_P(J^{-2\gamma-1} + T^{-1})$*, (iii)* $P(\max_{1\leq i\leq N}||\hat{Q}_{i,\tilde{z}\tilde{\eta}}|| > 2(\bar{c}_x)^{2/q}\vartheta_{NT}) = o(N^{-1})$ *where* $\vartheta_{NT} \equiv \max_{1\leq k\leq K}\sup_{v\in[0,1]}\|\alpha_k^0(v) - \omega_k^{0\prime}B(v)\| = O(J^{-\gamma})$*, and (iv)* $P(\max_{1\leq i\leq N}||\hat{Q}_{i,\tilde{z}\tilde{u}}|| \geq \epsilon T^{-1/2}(\ln T)^3)$ $= o(N^{-1})$ *for any* $\epsilon > 0$.

Lemma A.5 *Suppose that Assumption A1 holds. Then for any* $\epsilon > 0$ *(i)* $P(\max_{1\leq i\leq N} J^{-1/2}\|\tilde{\pi}_i - \pi_i^0\| > \epsilon) = o(N^{-1})$*, and (ii)* $P(\max_{1\leq i\leq N}\left|\tilde{\sigma}_i^2 - \bar{\sigma}_{i,T}^2\right| > \epsilon) = o(N^{-1})$.

Lemma A.6 *Suppose that Assumptions A1-A2 hold. Let* $\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} = \frac{1}{N_kT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$*, and* $\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} = \frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$. *Then (i)* $||J(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} - \bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})|| = o_P(J^{-1/2})$*;(ii) there exist finite positive constants* $\underline{c}_{zz}$ *and* $\bar{c}_{zz}$ *such that* $P(\underline{c}_{zz}/2 \leq \mu_{\min}(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}) \leq \mu_{\max}(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}) \leq 2\bar{c}_{zz}) = 1 - o(1)$.

Lemma A.7 *Suppose that Assumptions A1-A3 hold. Let* $\hat{Q}_{i,\tilde{z}\tilde{z}}^{(\sigma)} = \frac{1}{T}\sum_{t=1}^{T}\sigma_i^2(X_{it})\tilde{Z}_{it}\tilde{Z}_{it}'$ *and* $\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{(\sigma)} = \frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sigma_i^2(X_{it})\tilde{Z}_{it}\tilde{Z}_{it}'$. *Then there exist finite positive constants* $\underline{c}_{zz}^{(\sigma)}$ *and* $\bar{c}_{zz}^{(\sigma)}$ *such that (i)* $P(\max_{i\in G_k^0}\mu_{\max}(J\hat{Q}_{i,\tilde{z}\tilde{z}}^{(\sigma)}) \leq \bar{c}_{zz}^{(\sigma)}) = 1 - o(1)$*, and (ii)* $P(\underline{c}_{zz}^{(\sigma)} \leq \mu_{\min}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{(\sigma)}))$ $= 1 - o(1)$.

To state the next two lemmas, let $\mathbf{c}$ be an arbitrary nonrandom $p\times 1$ vector with $\|\mathbf{c}\| = 1$. Define $b_{\mathbf{c}} = \mathbf{c}\otimes B(v)$ and $\bar{b}_{\mathbf{c}} = b_{\mathbf{c}}/\|b_{\mathbf{c}}\|$. Let $\tilde{u}_{it} = u_{it} - \frac{1}{T}\sum_{t=1}^{T}u_{it}$ and $\tilde{\eta}_{it} = \eta_{it} - \frac{1}{T}\sum_{t=1}^{T}\eta_{it}$, where $\eta_{it} = \left[\beta_i^0(t/T) - \pi_i^0 B(t/T)\right]' X_{it}$.

Lemma A.8 *Suppose that Assumptions A1-A3 hold. Let* $A_{kNT} = b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{u}_{it}$ *and* $\hat{A}_{kNT} = b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{u}_{it}$. *Let* $S_{\mathbf{c},k}^2 = \mathbf{c}'\mathbb{S}_k\mathbf{c} = b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}$ $\times\{\frac{J}{N_kT}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}b_{\mathbf{c}}$ *denote the variance of* A_{kNT} *conditional on* $\mathcal{X} = \{X_1, ..., X_N\}$. *Then (i)* $S_{\mathbf{c},k} \asymp \|b_{\mathbf{c}}\| \asymp 1$*; (ii)* $A_{kNT}/S_{\mathbf{c},k} \xrightarrow{D} N(0,1)$*; and (iii)* $(\hat{A}_{kNT} - A_{kNT})/S_{\mathbf{c},k} = o_P(1)$.

Lemma A.9 *Suppose that Assumptions A1-A3 hold. Then* $\hat{C}_{kNT} = \bar{b}_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{\eta}_{it} = o_P(1)$.

Lemma A.10 *Suppose that Assumptions A1-A5 hold. Let* $\bar{\sigma}^2_{G^0} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}^2_{it}$. *Then* $\max_{K_0\le K\le K_{\max}}\left|\hat{\sigma}^2_{\hat{G}(K,\lambda)} - \bar{\sigma}^2_{G^0}\right| = O_P((NT/J)^{-1})$.

Proof of Theorem 4.1. (i) Recall that $\tilde{V}_i = \{\text{diag}(\frac{J}{T}\tilde{Z}_i'\tilde{Z}_i)\}^{1/2}$ and $\tilde{\sigma}^2_i = \frac{1}{T}\sum_{t=1}^{T}[\tilde{Y}_{it} - \tilde{Z}_{it}'\text{vec}(\tilde{\pi}_i)]^2$. By Lemma A.3(ii), $\tilde{V}_i$ is a diagonal matrix with diagonal elements bounded away from both zero and infinity uniformly in i w.p.a.1. By Lemma A.5 and Assumption A1(iv), $\tilde{\sigma}^2_i$'s are uniformly bounded away from zero by a positive constant (say, $\underline{c}_\sigma/2$) w.p.a.1.

Let $\hat{b}_i$ =vec$(\hat{\pi}_i - \pi^0_i)$ and b_i =vec$(\pi_i - \pi^0_i)$. Define $Q_{1,NT,i}(\pi_i) = \frac{1}{T}\sum_{t=1}^{T}[\tilde{Y}_{it} - \text{vec}(\pi_i)'\tilde{Z}_{it}]^2$ and $Q_{NT,i}(\pi_i,\boldsymbol{\omega}) = Q_{1,NT,i}(\pi_i) + \lambda\tilde{\sigma}_i^{2-K}\prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}(\pi_i - \omega_k)\right\|$. Then

$$Q_{1,NT,i}(\pi_i) - Q_{1,NT,i}(\pi^0_i) = b_i'\hat{Q}_{i,\tilde{z}\tilde{z}}b_i - 2b_i'\hat{Q}_{i,\tilde{z}\tilde{e}} \tag{A.2}$$

where $\hat{Q}_{i,\tilde{z}\tilde{z}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$ and $\hat{Q}_{i,\tilde{z}\tilde{e}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{e}_{it}$. By the triangle inequality, the fact that $\|Ab\| = \|Ab\|_{\text{sp}} \le \|A\|_{\text{sp}}\|b\|$ for conformable matrix A and vector b, and the fact that $\|\text{vec}(\pi_i)\| = \|\pi_i\|$, we have

$$\begin{aligned}
&\left|\prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_k)\right\| - \prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}\left(\pi^0_i - \omega_k\right)\right\|\right| \\
&\le \left|\prod_{k=1}^{K-1}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_k)\right\|\left\{\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_K)\right\| - \left\|\tilde{V}_i\text{vec}\left(\pi^0_i - \omega_K\right)\right\|\right\}\right| \\
&+ \left|\prod_{k=1}^{K-2}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_k)\right\|\left\|\tilde{V}_i\text{vec}(\pi_i - \omega_K)\right\|\left\{\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_{K-1})\right\| - \left\|\tilde{V}_i\text{vec}\left(\pi^0_i - \omega_{K-1}\right)\right\|\right\}\right| \\
&+ \ldots + \left|\prod_{k=2}^{K}\left\|\tilde{V}_i\text{vec}\left(\pi^0_i - \omega_k\right)\right\|\left\{\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_1)\right\| - \left\|\tilde{V}_i\text{vec}\left(\pi^0_i - \omega_1\right)\right\|\right\}\right| \\
&\le \hat{c}_{i,NT}(\boldsymbol{\omega})\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \pi^0_i\right)\right\| \le \hat{c}_{i,NT}(\boldsymbol{\omega})\left\|\tilde{V}_i\right\|_{\text{sp}}\left\|\hat{b}_i\right\|,
\end{aligned} \tag{A.3}$$

where $\hat{c}_{i,NT}(\boldsymbol{\omega}) = \prod_{k=1}^{K-1}||\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_k)|| + \prod_{k=1}^{K-2}||\tilde{V}_i\text{vec}(\hat{\pi}_i - \omega_k)||||\tilde{V}_i\text{vec}(\pi^0_i - \omega_K)|| + \ldots + \prod_{k=2}^{K}||\tilde{V}_i\ \text{vec}(\pi^0_i - \omega_k)|| = O_P(J^{(K-1)/2})$ as both π_i and ω_k are $J\times p$ dimensional matrices and $||\tilde{V}_i||_{\text{sp}} = O_P(1)$ by Lemma A.3(ii). Since $\hat{\pi}_i$ minimizes $Q_{NT,i}(\pi_i,\hat{\boldsymbol{\omega}})$, we have $Q_{NT,i}(\hat{\pi}_i,\hat{\boldsymbol{\omega}}) - Q_{NT,i}(\pi^0_i,\hat{\boldsymbol{\omega}}) \le 0$. This, in conjunction with the definition of $Q_{NT,i}$ and (A.2)-(A.3), implies that $\hat{b}_i'\hat{Q}_{i,\tilde{z}\tilde{z}}\hat{b}_i \le 2\hat{b}_i'\hat{Q}_{i,\tilde{z}\tilde{e}} + \lambda\hat{c}_{i,NT}(\hat{\boldsymbol{\omega}})||\tilde{V}_i||_{\text{sp}}||\hat{b}_i||$. Letting $\underline{\hat{c}}_{\tilde{z}\tilde{z}} = \min_{1\le i\le N}\mu_{\min}(J\hat{Q}_{i,\tilde{z}\tilde{z}})$, we have $\underline{\hat{c}}_{\tilde{z}\tilde{z}}||\hat{b}_i||^2 \le J[2||\hat{Q}_{i,\tilde{z}\tilde{e}}|| + \lambda\hat{c}_{i,NT}(\hat{\boldsymbol{\omega}})||\tilde{V}_i||_{\text{sp}}]||\hat{b}_i||$, or equivalently,

$$\left\|\hat{b}_i\right\| \le \underline{\hat{c}}_{\tilde{z}\tilde{z}}^{-1}J\left[2\left\|\hat{Q}_{i,\tilde{z}\tilde{e}}\right\| + \lambda\hat{c}_{i,NT}(\hat{\boldsymbol{\omega}})\left\|\tilde{V}_i\right\|_{\text{sp}}\right]. \tag{A.4}$$

Then by Lemmas A.3 and A.4(i),

$$\left\|\hat{b}_i\right\| = J[O_P(J^{-\gamma-1/2} + T^{-1/2}) + \lambda J^{(K-1)/2}] = O_P(J^{-\gamma+1/2} + JT^{-1/2} + \lambda J^{(K+1)/2}). \tag{A.5}$$

(ii) By Minkowski's inequality, we have:

$$\begin{aligned}
\hat{c}_{i,NT}(\boldsymbol{\omega}) &\leq \prod_{k=1}^{K-1}\left[\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\pi_i^0\right)\right\|+\left\|\tilde{V}_i\text{vec}\left(\pi_i^0-\omega_k\right)\right\|\right] \\
&+\prod_{k=1}^{K-2}\left[\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\pi_i^0\right)\right\|+\left\|\tilde{V}_i\text{vec}\left(\pi_i^0-\omega_k\right)\right\|\right]\left\|\tilde{V}_i\text{vec}\left(\pi_i^0-\omega_K\right)\right\| \\
&+\ldots+\prod_{k=2}^{K}\left\|\tilde{V}_i\text{vec}\left(\pi_i^0-\omega_k\right)\right\| \\
&=\sum_{s=0}^{K-1}\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\pi_i^0\right)\right\|^s\prod_{k=1}^{s}a_{ks}\left\|\tilde{V}_i\text{vec}\left(\pi_i^0-\omega_k\right)\right\|^{K-1-s} \\
&\leq C_{K,NT}(\boldsymbol{\omega})\sum_{s=0}^{K-1}\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\pi_i^0\right)\right\|^s \\
&\leq C_{K,NT}(\boldsymbol{\omega})\left(1+2\left\|\tilde{V}_i\right\|_{\text{sp}}\left\|\hat{\pi}_i-\pi_i^0\right\|\right) \qquad \text{(A.6)}
\end{aligned}$$

where a_{ks}'s are finite integers and $C_{K,NT}(\boldsymbol{\omega})=\max_{1\leq i\leq N}\max_{0\leq s\leq k\leq K-1}\prod_{k=1}^{s}a_{ks}||\tilde{V}_i\text{vec}(\pi_i^0$ $-\omega_k)||^{K-1-s}=\max_{1\leq l\leq K}\max_{0\leq s\leq k\leq K-1}\prod_{k=1}^{s}a_{ks}|||\tilde{V}_i\text{vec}(\omega_l^0-\omega_k)||^{K-1-s}=O(J^{(K-1)/2})$ as ω_k is a $J\times p$ matrix. Let $\varsigma_{NT}=2C_{K,NT}(\hat{\boldsymbol{\omega}})||\tilde{V}_i||_{\text{sp}}\lambda J\underline{\hat{c}}_{\tilde{z}\tilde{z}}^{-1}$. In view of the fact that $\varsigma_{NT}=O_P\left(\lambda J^{(K+1)/2}\right)=o_P(1)$, we can combine (A.4) with (A.6) to yield

$$\left\|\hat{b}_i\right\|\leq\frac{\underline{\hat{c}}_{\tilde{z}\tilde{z}}^{-1}}{1-\varsigma_{NT}}J\left[2\left\|\hat{Q}_{i,\tilde{z}\tilde{e}}\right\|+\lambda C_{K,NT}(\hat{\boldsymbol{\omega}})\left\|\tilde{V}_i\right\|_{\text{sp}}\right].$$

Then by Lemmas A.3-A.4

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2 &\leq \left(\frac{\underline{\hat{c}}_{\tilde{z}\tilde{z}}^{-1}}{1-\varsigma_{NT}}\right)^2\frac{J^2}{N}\sum_{i=1}^{N}\left[8\left\|\hat{Q}_{i,\tilde{z}\tilde{e}}\right\|^2+2\lambda^2C_{K,NT}(\hat{\boldsymbol{\omega}})^2\left\|\tilde{V}_i\right\|_{\text{sp}}^2\right] \\
&= J^2O_P(J^{-2\gamma-1}+T^{-1}+\lambda^2J^{K-1})=O_P(J^{-2\gamma+1}+J^2T^{-1}+\lambda^2J^{K+1}). \qquad \text{(A.7)}
\end{aligned}$$

To refine the result in (A.7), we need to demonstrate $\frac{1}{N}\sum_{i=1}^{N}||\hat{b}_i||^2=O_P\left(c_{NT}^2\right)$ where $c_{NT}=J^{-\gamma+1/2}+JT^{-1/2}$. Recall that $\boldsymbol{\pi}=(\text{vec}(\pi_1)',\ldots,\text{vec}(\pi_N)')'$. Let $\boldsymbol{\pi}=\boldsymbol{\pi}^0+c_{NT}\boldsymbol{\nu}$ where $\boldsymbol{\nu}=(\text{vec}(\nu_1)',\ldots,\text{vec}(\nu_N)')$ with ν_i's being $J\times p$ matrices. We want to show that for any given $\epsilon^*>0$, there exists a large constant $M=M\left(\epsilon^*\right)$ such that, for sufficiently large N and T we have

$$P\left\{\inf_{N^{-1}\sum_{i=1}^{N}\|\nu_i\|^2=M}Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0+c_{NT}\boldsymbol{\nu},\hat{\boldsymbol{\omega}}\right)>Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0,\boldsymbol{\omega}^0\right)\right\}\geq 1-\epsilon^*. \qquad \text{(A.8)}$$

This implies that w.p.a.1 there is a local minimum $\{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\omega}}\}$ such that $N^{-1}\sum_{i=1}^{N}||\hat{b}_i||^2 = O_P(c_{NT}^2)$ regardless of the property of $\hat{\boldsymbol{\omega}}$. By (A.2) and the Cauchy-Schwarz inequality

$$
\begin{aligned}
& Jc_{NT}^{-2}\left[Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0 + c_{NT}\boldsymbol{\nu}, \hat{\boldsymbol{\omega}}\right) - Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0, \boldsymbol{\omega}^0\right)\right] \\
= & \frac{1}{N}\sum_{i=1}^{N}\text{vec}(\nu_i)'\left(J\hat{Q}_{i,\tilde{z}\tilde{z}}\right)\text{vec}(\nu_i) - \frac{2Jc_{NT}^{-1}}{N}\sum_{i=1}^{N}\text{vec}(\nu_i)'\hat{Q}_{i,\tilde{z}\tilde{e}} \\
& + \frac{\lambda J\delta_{NT}^{-2}}{N}\sum_{i=1}^{N}\tilde{\sigma}_{e,i}^{2-K}\prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}\left(\pi_i^0 + c_{NT}\nu_i - \hat{\omega}_k\right)\right\| \\
\geq & \underline{\hat{c}}_{\tilde{z}\tilde{z}}\frac{1}{N}\sum_{i=1}^{N}\|v_i\|^2 - 2\left\{\frac{1}{N}\sum_{i=1}^{N}\|v_i\|^2\right\}^{1/2}\left\{\frac{J^2c_{NT}^{-2}}{N}\sum_{i=1}^{N}\left\|\hat{Q}_{i,\tilde{z}\tilde{e}}\right\|^2\right\}^{1/2} \\
\equiv & D_{1NT} - D_{2NT}, \text{ say.}
\end{aligned}
$$

By Lemma A.3(i), $\underline{\hat{c}}_{\tilde{z}\tilde{z}}$ is bounded below by $\underline{c}_{\tilde{z}\tilde{z}} > 0$ w.p.a.1. By Lemma A.4(ii), $\frac{J^2c_{NT}^{-2}}{N}\sum_{i=1}^{N}||\hat{Q}_{i,\tilde{z}\tilde{e}}||^2 = J^2c_{NT}^{-2}O_P\left(J^{-2\gamma-1} + T^{-1}\right) = O_P(1)$. So D_{1NT} dominates D_{2NT} for sufficiently large M. That is $Jc_{NT}^{-2}[Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0 + c_{NT}\boldsymbol{\nu}, \hat{\boldsymbol{\omega}}\right) - Q_{NT,\lambda}^{(K)}\left(\boldsymbol{\pi}^0, \boldsymbol{\omega}^0\right)] > 0$ for sufficiently large M and we cannot achieve minimization. Consequently, we must have $N^{-1}\sum_{i=1}^{N}||\hat{b}_i||^2 = O_P\left(c_{NT}^2\right)$.

(iii) Let $P_{NT}(\boldsymbol{\pi}, \boldsymbol{\omega}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\prod_{k=1}^{K}\|\tilde{V}_i\text{vec}(\pi_i - \omega_k)\|$. Observe that

$$
\begin{aligned}
0 &\geq P_{NT}(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\omega}}) - P_{NT}(\hat{\boldsymbol{\pi}}, \boldsymbol{\omega}^0) \\
&= \left[P_{NT}(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\omega}}) - P_{NT}(\boldsymbol{\pi}^0, \hat{\boldsymbol{\omega}})\right] + \left[P_{NT}(\boldsymbol{\pi}^0, \hat{\boldsymbol{\omega}}) - P_{NT}(\boldsymbol{\pi}^0, \boldsymbol{\omega}^0)\right] \\
&\quad - \left[P_{NT}(\hat{\boldsymbol{\pi}}, \boldsymbol{\omega}^0) - P_{NT}(\boldsymbol{\pi}^0, \boldsymbol{\omega}^0)\right]. \qquad \text{(A.9)}
\end{aligned}
$$

By (A.3), (A.6), and (A.7),

$$
\begin{aligned}
|P_{NT}(\hat{\boldsymbol{\pi}}, \boldsymbol{\omega}) - P_{NT}(\boldsymbol{\pi}, \boldsymbol{\omega})| &\leq C_{K,NT}(\boldsymbol{\omega})\frac{1}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\left\{\left\|\tilde{V}_i\right\|_{\text{sp}}\|\hat{b}_i\| + 2\left\|\tilde{V}_i\right\|_{\text{sp}}^2\|\hat{b}_i\|^2\right\} \\
&\leq C_{K,NT}(\boldsymbol{\omega})c_{1NT}\left\{c_{2NT}\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2\right)^{1/2} + \frac{2c_{3NT}}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2\right\} \\
&= O(J^{(K-1)/2})O_P(1)\,O_P(J^{-\gamma+1/2} + JT^{-1/2}) \\
&= O_P(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2}), \qquad \text{(A.10)}
\end{aligned}
$$

where $c_{1NT} = \max_{1\leq i\leq N}\tilde{\sigma}_i^{2-K}$, $c_{2NT} = \{\frac{1}{N}\sum_{i=1}^{N}||\tilde{V}_i||_{\text{sp}}^2\}^{1/2}$, and $c_{3NT} = \max_{1\leq i\leq N}||\tilde{V}_i||_{\text{sp}}^2$, all of which are $O_P(1)$ by the remark at the beginning of the proof. In addition,

$$
\begin{aligned}
P_{NT}(\boldsymbol{\pi}^0, \hat{\boldsymbol{\omega}}) - P_{NT}(\boldsymbol{\pi}^0, \boldsymbol{\omega}^0) &= \frac{1}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\prod_{k=1}^{K}\left\|\tilde{V}_i\text{vec}(\hat{\omega}_k - \pi_i^0)\right\| \\
&\geq \frac{\varsigma_{1NT}N_1}{N}\prod_{k=1}^{K}\left\|\hat{\omega}_k - \omega_1^0\right\| + ... + \frac{\varsigma_{KNT}N_K}{N}\prod_{k=1}^{K}\left\|\hat{\omega}_k - \omega_K^0\right\| \qquad \text{(A.11)}
\end{aligned}
$$

where $\varsigma_{kNT} = (\min_{i\in G_k^0} \tilde{\sigma}_i^{2-K}) \min_{i\in G_k^0} \mu_{\min}(\tilde{V}_i)$ for $k = 1, ..., K$ and we have used the fact $\|AB\|^2 = \mathrm{tr}(BB'A^2) \geq \mu_{\min}(A^2)\mathrm{tr}(BB')$ for symmetric matrix A and conformable matrix B (e.g., Bernstein 2005, p.275). Combining (A.9)-(A.11), we have

$$0 \geq \frac{\varsigma_{1NT}N_1}{N}\prod_{k=1}^{K}\left\|\hat{\omega}_k - \omega_1^0\right\| + ... + \frac{\varsigma_{KNT}N_K}{N}\prod_{k=1}^{K}\left\|\hat{\omega}_k - \omega_K^0\right\| + O_P\left(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2}\right),$$

which, by Assumption A1(vi) and the fact that ς_{kNT} is bounded away from zero, implies that $\prod_{k=1}^{K}\|\hat{\omega}_k - \omega_l^0\| = O_P\left(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2}\right)$ for $l = 1, ..., K$. This further implies that there is a permutation $\left(\hat{\omega}_{(1)}, .., \hat{\omega}_{(K)}\right)$ of $(\hat{\omega}_1, .., \hat{\omega}_K)$ such that $\left\|\hat{\omega}_{(k)} - \omega_k^0\right\| = \min_{1\leq l\leq K}\|\hat{\omega}_l - \omega_k^0\|$ and $\left\|\hat{\omega}_{(k)} - \omega_l^0\right\| \geq \|\omega_k^0 - \omega_l^0\| - \left\|\hat{\omega}_{(k)} - \omega_k^0\right\| \asymp \|\omega_k^0 - \omega_l^0\| \asymp J^{1/2}$ for $l \neq k$ by (4.3) as we can easily show that $\left\|\hat{\omega}_{(k)} - \omega_k^0\right\|$ must be smaller than $J^{1/2}$ in probability order by contradiction under Assumption A1(v). Consequently, $\left\|\hat{\omega}_{(k)} - \omega_k^0\right\| = J^{-(K-1)/2}O_P(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2}) = O_P(J^{-\gamma+1/2} + JT^{-1/2})$ for each $k = 1, ..., K$. ■

Proof of Corollary 4.2. We only prove (i) as the proof of (ii) is analogous. First, we make the decomposition: $\hat{\beta}_i(v) - \beta_i^0(v) = [\hat{\pi}_i - \pi_i^0]' B(v) + \pi_i^{0\prime} B(v) - \beta_i^0(v)$. By Properties (i)-(ii) of B-splines, $\|B(v)\| = \{\sum_{j=-d+1}^{J_0} B_j(v)^2\}^{1/2} \leq \{\sum_{j=-d+1}^{J_0} B_j(v)\}^{1/2} = 1$. This, in conjunction with Theorem 4.1(i), implies that,

$$\sup_{v\in[0,1]}\left\|\left[\hat{\pi}_i - \pi_i^0\right]' B(v)\right\| \leq \left\|\hat{\pi}_i - \pi_i^0\right\| = O_P(J^{-\gamma+1/2} + JT^{-1/2} + \lambda J^{(K+1)/2}).$$

By Assumption A1(v), $\sup_{v\in[0,1]}\left\|\pi_i^{0\prime}B(v) - \beta_i^0(v)\right\| = O_P(J^{-\gamma})$. Consequently, $\sup_{v\in[0,1]} ||\hat{\beta}_i(v) - \beta_i^0(v)|| = O_P(J^{-\gamma+1/2} + JT^{-1/2} + \lambda J^{(K+1)/2})$.

Note that $\int_0^1 \left\|\hat{\beta}_i(v) - \beta_i^0(v)\right\|^2 dv \leq 2\int \left\|[\hat{\pi}_i - \pi_i^0]' B(v)\right\|^2 dv + 2\int_0^1 ||\pi_i^{0\prime}B(v) - \beta_i^0(v)||^2 dv \equiv 2A_{1i} + 2A_{2i}$, say. By Property (iii) of B-splines and Theorem 4.1(i),

$$\begin{aligned} A_{1i} &= 2\int_0^1 \left\|\left(\hat{\pi}_i - \pi_i^0\right)' B(v)\right\|^2 dv = 2\mathrm{tr}\left\{\left(\hat{\pi}_i - \pi_i^0\right)'\left(\int_0^1 B(v)B(v)' dv\right)\left(\hat{\pi}_i - \pi_i^0\right)\right\} \\ &\asymp J^{-1}\left\|\hat{b}_i\right\|^2 = O_P\left(J^{-2\gamma} + JT^{-1} + \lambda^2 L^K\right). \end{aligned}$$

By Assumption A1(v), $A_{2i} = O(J^{-2\gamma})$. Thus the result in (i) follows. ■

Proof of Theorem 4.3. (i) Fix $k \in \{1, 2, ..., K\}$. By Theorem 4.1, $||\hat{\beta}_i - \hat{\omega}_l|| \xrightarrow{P} \|\omega_k^0 - \omega_l^0\| \neq 0$ for all $i \in G_k^0$ and $l \neq k$. Let $\hat{c}_{ki} \equiv \prod_{l=1,l\neq k}^{K} \|\tilde{V}_i\mathrm{vec}(\hat{\pi}_i - \hat{\omega}_l)\|$ and $c_{ki}^0 = \prod_{l=1,l\neq k}^{K} \|V_i^0(\omega_k^0 - \omega_l^0)\|$, where V_i^0 denotes the probability limit of $\tilde{V}_i$. Note that

$$\begin{aligned} \hat{c}_{ki} &= \prod_{l=1,l\neq k}^{K} \|\tilde{V}_i\mathrm{vec}\left\{\left(\hat{\pi}_i - \omega_k^0\right) - \left(\hat{\omega}_l - \omega_l^0\right) + \left(\omega_k^0 - \omega_l^0\right)\right\}\| \\ &= \prod_{l=1,l\neq k}^{K} \{||V_i^0\mathrm{vec}\left(\omega_k^0 - \omega_l^0\right)|| + o_P(1)\} \asymp c_{ki}^0 \asymp J^{(K-1)/2} \text{ for any } i \in G_k^0. \end{aligned}$$

Now, suppose that $\|\hat{\pi}_i - \hat{\omega}_k\| \neq 0$ for some $i \in G_k^0$. Then the first order condition with respect to $\text{vec}(\pi_i)$ for the minimization problem in (3.4) implies that:

$$
\begin{aligned}
\mathbf{0}_{Jp\times 1} &= \frac{-2}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\left[\tilde{Y}_{it} - \tilde{Z}_{it}'\text{vec}(\hat{\pi}_i)\right] + \lambda\sum_{j=1}^{K}\frac{\tilde{V}_i^2\text{vec}(\hat{\pi}_i - \hat{\omega}_j)}{\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \hat{\omega}_j\right)\|}\prod_{l=1,l\neq j}^{K}\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \hat{\omega}_l\right)\| \\
&= -2\hat{Q}_{i,\tilde{z}\tilde{u}} + \left[2\hat{Q}_{i,\tilde{z}\tilde{z}}\tilde{V}_i^{-1} + \frac{\lambda\hat{c}_{ki}}{\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \hat{\omega}_k\right)\|}\tilde{V}_i\right]\left[\tilde{V}_i\text{vec}(\hat{\pi}_i - \hat{\omega}_k)\right] - 2\hat{Q}_{i,\tilde{z}\tilde{\eta}} \\
&\quad + 2\hat{Q}_{i,\tilde{z}\tilde{z}}\text{vec}(\hat{\omega}_k - \pi_i^0) + \lambda\sum_{j=1,j\neq k}^{K}\frac{\tilde{V}_i^2\text{vec}(\hat{\pi}_i - \hat{\omega}_j)}{\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \hat{\omega}_j\right)\|}\prod_{l=1,l\neq j}^{K}\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i - \hat{\omega}_l\right)\| \\
&\equiv \hat{B}_{i1} + \hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5},\ \text{say,}
\end{aligned}
$$

where $\hat{Q}_{i,\tilde{z}\tilde{z}}$ is defined in Lemma A.3, and $\hat{Q}_{i,\tilde{z}\tilde{u}}$ and $\hat{Q}_{i,\tilde{z}\tilde{\eta}}$ are defined in Lemma A.4.

Let $\varkappa_{NT} = J^{-\gamma+1/2}\left(\ln T\right)^{\nu} + JT^{-1/2}\left(\ln T\right)^3 + \lambda J^{(K+1)/2}\left(\ln T\right)^{\nu}$, $\bar{c}_k^0 = \max_{i\in G_k^0} c_{ki}^0$, $\underline{c}_k^0 = \min_{i\in G_k^0} c_{ki}^0$, and $\hat{\vartheta}_{ki} = [\tilde{V}_i\text{vec}(\hat{\pi}_i - \hat{\omega}_k)]/||\tilde{V}_i\text{vec}(\hat{\pi}_i - \hat{\omega}_k)||$. Let C denote a generic constant that may vary across lines. Following the proof of Theorem 4.1 and using the uniform results in Lemmas A.3–A.5, we can show that

$$P\left(\max_i\left\|\hat{\pi}_i - \pi_i^0\right\| \geq C\varkappa_{NT}\right) = o\left(N^{-1}\right)\ \text{for some } C > 0, \tag{A.12}$$

$$P\left(\left\|\hat{\omega}_k - \omega_k^0\right\| \geq C\left(J^{-\gamma+1/2} + JT^{-1/2}\right)\left(\ln T\right)^{\nu}\right) = o\left(N^{-1}\right), \tag{A.13}$$

$$P\left(\underline{c}_k^0/2 \leq \hat{c}_{ki} \leq 2\bar{c}_k^0\right) = 1 - o\left(N^{-1}\right). \tag{A.14}$$

By (A.12)–(A.14) and Lemma A.3, $P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\| \geq C\lambda\varkappa_{NT}\right) = o\left(N^{-1}\right)$ for some $C > 0$. Combining these results with those in Lemmas A.3-A.4, we have $P\left(\Xi_{kNT}\right) = 1 - o\left(N^{-1}\right)$, where

$$
\begin{aligned}
\Xi_{kNT} \equiv\ & \left\{\underline{c}_k^0/2 \leq \hat{c}_{ki} \leq 2\bar{c}_k^0\right\} \cap \left\{\left\|\hat{\omega}_k - \omega_k^0\right\| \leq C\left(J^{-\gamma+1/2} + JT^{-1/2}\right)\left(\ln T\right)^{\nu}\right\} \\
& \cap\left\{\underline{c}_{zz} \leq \min_{1\leq i\leq N}\mu_{\min}(J\hat{Q}_{i,\tilde{z}\tilde{z}}) \leq \max_{1\leq i\leq N}\mu_{\max}(J\hat{Q}_{i,\tilde{z}\tilde{z}}) \leq \bar{c}_{zz}\right\} \\
& \cap\left\{\underline{c}_{zz}^{1/2} \leq \min_{1\leq i\leq N}\mu_{\min}(\tilde{V}_i) \leq \max_{1\leq i\leq N}\mu_{\max}(\tilde{V}_i) \leq \bar{c}_{zz}^{1/2}\right\} \\
& \cap\left\{\max_{1\leq i\leq N}\left\|\hat{Q}_{i,\tilde{z}\tilde{\eta}}\right\| \leq 2\left(\bar{c}_x\right)^{2/q}\vartheta_{NT}\right\},
\end{aligned}
$$

where $\vartheta_{NT} \equiv \max_{1\leq k\leq K}\sup_{v\in[0,1]}\|\alpha_k^0\left(v\right) - \omega_k^{0\prime}B\left(v\right)\| = O\left(J^{-\gamma}\right)$. Then conditional on Ξ_{kNT}, we have uniformly in $i \in G_k^0$,

$$\hat{\vartheta}_{ki}'\hat{B}_{i2} \geq \lambda \hat{c}_{ki}\hat{\vartheta}_{ik}'\tilde{V}_i\hat{\vartheta}_{ik} \geq \lambda \hat{c}_{ki}\mu_{\min}(\tilde{V}_i) \geq \lambda \underline{c}_{zz}^{1/2}\underline{c}_k^0/2,$$
$$\left|\hat{\vartheta}_{ki}'\hat{B}_{i3}\right| \leq 2\left\|\hat{Q}_{i,\tilde{z}\tilde{\eta}}\right\| \leq 4\left(\bar{c}_x\right)^{2/q}\vartheta_{NT},$$
$$\left|\hat{\vartheta}_{ki}'\hat{B}_{i4}\right| \leq 2J^{-1}\left\|J\hat{Q}_{i,\tilde{z}\tilde{z}}\right\|_{\mathrm{sp}}\left\|\hat{\omega}_k - \omega_k^0\right\| \leq 2CJ^{-1}\bar{c}_{zz}\left(J^{-\gamma+1/2} + JT^{-1/2}\right)(\ln T)^{\nu},$$
$$\left|\hat{\vartheta}_{ki}'\hat{B}_{i5}\right| \leq \max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\| \leq C\lambda\varkappa_{NT},$$

and

$$\begin{aligned} & \left|\hat{\vartheta}_{ki}'\left(\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right| \\ \geq\ & \left|\hat{\vartheta}_{ki}'\hat{B}_{i2}\right| - \left|\hat{\vartheta}_{ki}'\left(\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right| \\ \geq\ & \lambda \underline{c}_{zz}^{1/2}\underline{c}_k^0/2 - \left[4\left(\bar{c}_x\right)^{2/q}\vartheta_{NT} + 2CJ^{-1}\bar{c}_{zz}\left(J^{-\gamma+1/2}+JT^{-1/2}\right)(\ln T)^{\nu} + C\lambda\varkappa_{NT}\right] \\ \geq\ & \lambda \underline{c}_{zz}^{1/2}\underline{c}_k^0/4 \text{ for sufficiently large } (N,T), \end{aligned}$$

where the last inequality follows because $\lambda J^{(K-1)/2} \gg J^{-\gamma}+J^{-1}\left(J^{-\gamma+1/2}+JT^{-1/2}\right)(\ln T)^{\nu}+\lambda\varkappa_{NT}$ by Assumption A2. Then for all $i \in G_k^0$ we have

$$\begin{aligned} P(\hat{E}_{kNT,i}) &= P\left(i \notin \hat{G}_k \mid i \in G_k^0\right) = P\left(-\hat{B}_{i1} = \hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right) \\ &\leq P\left(\left|\hat{\vartheta}_{ki}'\hat{B}_{i1}\right| \geq \left|\hat{\vartheta}_{ki}'\left(\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right|\right) \\ &\leq P\left(\left\|\hat{B}_{i1}\right\| \geq \lambda\underline{c}_{zz}^{1/2}\underline{c}_k^0/4,\ \Xi_{kNT}\right) + P\left(\Xi_{kNT}^c\right) = o\left(N^{-1}\right), \end{aligned}$$

where Ξ_{kNT}^c denotes the complement of Ξ_{kNT} and the convergence follows by Lemma A.4 and Assumption A2 . Consequently, we can conclude that with probability $1-o(N^{-1})$ the differences $\hat{\pi}_i - \hat{\omega}_k$ must reach the point where $\|\tilde{V}_i \mathrm{vec}(\pi_i - \omega_k)\|$ is not differentiable with respect to $\mathrm{vec}(\pi_i)$ for some $i \in G_k^0$. That is, $P\left(\|\hat{\pi}_i - \hat{\omega}_k\| = 0 | i \in G_k^0\right) = 1 - o\left(N^{-1}\right)$.

For uniform consistency, we have: $P(\cup_{k=1}^{K_0}\hat{E}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0} P(\hat{E}_{kNT,i})$ and by Lemma A.4(iv) and Assumption A2(ii),

$$\begin{aligned} \sum_{k=1}^{K_0}\sum_{i\in G_k^0} P\left(\hat{E}_{kNT,i}\right) &\leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0}\left[P\left(\left\|\hat{B}_{i1}\right\| \geq \lambda\underline{c}_{zz}^{1/2}\underline{c}_k^0/4,\ \Xi_{kNT}\right) + P\left(\Xi_{kNT}^c\right)\right] \\ &\leq N\max_{1\leq i\leq N} P\left(\left\|\hat{Q}_{i,\tilde{z}\tilde{u}}\right\| \geq \lambda\underline{c}_{zz}^{1/2}\underline{c}_k^0/4\right) + o(1) = o(1). \end{aligned} \tag{A.15}$$

This completes the proof of Theorem 4.3(i).

(ii) Given (i), the proof is identical to Theorem 2(ii) in SSP and thus omitted. ■

Proof of Theorem 4.4. (i) We first show that $\sqrt{NT/J}\left[\hat{\alpha}_k(v)-\hat{\alpha}_{G_k}(v)\right]=o_P(1)$. Noting that $\sqrt{NT/J}\left\|\hat{\alpha}_k\left(v\right)-\hat{\alpha}_{G_k}\left(v\right)\right\| \leq \sqrt{NT/J}\left\|\hat{\omega}_k-\hat{\omega}_{G_k}\right\|\left\|B\left(v\right)\right\|$ and $\left\|B\left(v\right)\right\|^2=\sum_{j=1}^{J}B_j\left(v\right)^2 \leq \sum_{j=1}^{J}B_j\left(v\right)=1$ by Properties (i)-(ii) of B-splines and Lemma A.8(i), we can prove (i) by showing that $\sqrt{NT/J}\left\|\hat{\omega}_k-\hat{\omega}_{G_k}\right\|=o_P(1)$.

Based on subdifferential calculus (e.g., Bertsekas 1995, Appendix B.5), the first order conditions for the minimization of the objective function $Q_{NT,\lambda}^{(K)}(\boldsymbol{\pi},\boldsymbol{\omega})$ in (3.4) with respect to π_i and ω_k are given by

$$\mathbf{0}_{Jp\times 1}=\frac{-2}{NT}\sum_{t=1}^{T}\tilde{Z}_{it}\left[\tilde{Y}_{it}-\tilde{Z}_{it}'\mathrm{vec}(\hat{\pi}_i)\right]+\frac{\lambda}{N}\sum_{j=1}^{K}\tilde{\sigma}_i^{2-K}\hat{c}_{ij}\prod_{l=1,l\neq j}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|, \tag{A.16}$$

$$\mathbf{0}_{Jp\times 1}=\frac{\lambda}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|, \tag{A.17}$$

where $\hat{c}_{ij}=\frac{\tilde{V}_i^2\mathrm{vec}(\hat{\pi}_i-\hat{\omega}_j)}{\left\|\tilde{V}_i\mathrm{vec}(\hat{\pi}_i-\hat{\omega}_j)\right\|}$ if $\left\|\hat{\pi}_i-\hat{\omega}_j\right\|\neq 0$ and $\left\|\hat{c}_{ij}\right\|\leq ||\tilde{V}_i||_{\mathrm{sp}}$ if $\left\|\hat{\pi}_i-\hat{\omega}_j\right\|=0$. Fix $k\in\{1,2,...,K\}$. Then (a) $\left\|\hat{\pi}_i-\hat{\omega}_k\right\|=0$ for any $i\in\hat{G}_k$ by the definition of $\hat{G}_k$; (b) $\left\|\hat{\pi}_i-\hat{\omega}_l\right\|=\left\|(\hat{\pi}_i-\omega_k^0)-(\hat{\omega}_l-\omega_l^0)+(\omega_k^0-\omega_k^0)\right\|\geq\left\|\omega_k^0-\omega_l^0\right\|-o_P\left(1\right)\asymp\left\|\omega_k^0-\omega_l^0\right\|$ for any $i\in\hat{G}_k$ and $l\neq k$ by Theorem 4.1. It follows that $\left\|\hat{c}_{ik}\right\|\leq||\tilde{V}_i||_{\mathrm{sp}}$ for any $i\in\hat{G}_k$ and $\hat{c}_{ij}=\frac{\tilde{V}_i^2\mathrm{vec}(\hat{\pi}_i-\hat{\omega}_j)}{\left\|\tilde{V}_i\mathrm{vec}(\hat{\pi}_i-\hat{\omega}_j)\right\|}=\frac{\tilde{V}_i^2\mathrm{vec}(\hat{\omega}_k-\hat{\omega}_j)}{\left\|\tilde{V}_i\mathrm{vec}(\hat{\omega}_k-\hat{\omega}_j)\right\|}$ w.p.a.1 for any $i\in\hat{G}_k$ and $j\neq k$. Let $\hat{G}_0$ denote the set of unclassified individuals. As a result, we have that w.p.a.1

$$\begin{aligned}&\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^{K}\tilde{\sigma}_i^{2-K}\hat{c}_{ij}\prod_{l=1,l\neq j}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|\\ =&\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^{K}\tilde{\sigma}_i^{2-K}\frac{\tilde{V}_i^2\mathrm{vec}\left(\hat{\omega}_k-\hat{\omega}_j\right)}{\left\|\tilde{V}_i\left(\hat{\omega}_k-\hat{\omega}_j\right)\right\|}\prod_{l=1,l\neq j}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\omega}_k-\hat{\omega}_l\right)\right\|=\mathbf{0}_{Jp\times 1},\end{aligned} \tag{A.18}$$

and

$$\begin{aligned}&\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|\\ &=\sum_{i\in\hat{G}_k}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\omega}_k-\hat{\omega}_l\right)\right\|+\sum_{i\in\hat{G}_0}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|\\ &+\sum_{j=1,j\neq k}^{K}\sum_{i\in\hat{G}_j}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\omega}_j-\hat{\omega}_l\right)\right\|\\ &=\sum_{i\in\hat{G}_k}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\omega}_k-\hat{\omega}_l\right)\right\|+\sum_{i\in\hat{G}_0}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\mathrm{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|\\ &=\mathbf{0}_{Jp\times 1}.\end{aligned} \tag{A.19}$$

Combine (A.18) and (A.19) with (A.16), we have

$$\frac{2}{NT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\left[\tilde{Y}_{it}-\tilde{Z}_{it}'\text{vec}(\hat{\pi}_i)\right]+\frac{\lambda}{N}\sum_{i\in\hat{G}_0}\tilde{\sigma}_i^{2-K}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|=\mathbf{0}_{Jp\times 1}.$$

It follows that

$$\begin{aligned}\text{vec}(\hat{\omega}_k)&=\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{-1}\frac{1}{NT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Y}_{it}+\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{-1}\frac{\lambda}{2N}\sum_{i\in\hat{G}_0}\hat{c}_{ik}\prod_{l=1,l\neq k}^{K}\left\|\tilde{V}_i\text{vec}\left(\hat{\pi}_i-\hat{\omega}_l\right)\right\|\\&\equiv\text{vec}(\hat{\omega}_{\hat{G}_k})+\hat{R}_k,\ \text{say.}\end{aligned}$$

where $\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}=\frac{1}{N_kT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$. Thus, for any $\epsilon>0$, we have by Theorem 4.3(i)

$$\begin{aligned}P\left(\sqrt{NT/J}\left\|\tilde{V}_i\text{vec}\left(\hat{\omega}_k-\hat{\omega}_{\hat{G}_k}\right)\right\|\geq\epsilon\right)&=P\left(\sqrt{NT/J}\left\|\hat{R}_k\right\|\geq\epsilon\right)\leq\sum_{k=1}^{K}\sum_{i\in G_k}P\left(i\in\hat{G}_0|i\in G_k^0\right)\\&\leq\sum_{k=1}^{K}\sum_{i\in G_k}P\left(i\notin\hat{G}_k|i\in G_k^0\right)=o(1).\end{aligned}$$

Consequently, we have $\sqrt{NT/J}||\tilde{V}_i\text{vec}(\hat{\omega}_k-\hat{\omega}_{\hat{G}_k})||=o_P(1)$, implying that $\sqrt{NT/J}||\hat{\omega}_k-\hat{\omega}_{\hat{G}_k}||=o_P(1)$ by Lemma A.3(ii). It follows that $\sqrt{NT/L}[\hat{\alpha}_k(v)-\hat{\alpha}_{\hat{G}_k}(v)]=o_P(1)$. Then (i) holds by result in part (ii).

(ii) We first make the following decomposition:

$$\begin{aligned}\sqrt{N_kT/J}\left[\hat{\alpha}_{\hat{G}_k}(v)-\alpha_k^0(v)\right]&=\sqrt{N_kT/J}\left(\hat{\omega}_{\hat{G}_k}-\omega_k^0\right)'B(v)+\sqrt{N_kT/J}\left[\omega_k^{0\prime}B(v)-\alpha_k^0(v)\right]\\&\equiv\mathbb{D}_{1k}+\mathbb{D}_{2k},\ \text{say.}\end{aligned}$$

By Assumptions A1(v) and A3(iv) and Lemma A.8(i), $\mathbb{D}_{2k}/S_{\mathbf{c},k}=\sqrt{N_kTJ^{-1}}O\left(J^{-\gamma}\right)=O(\sqrt{N_kTJ^{-1-2\gamma}})=o(1)$. It suffices to prove (ii) by showing that $\mathbb{D}_{1k,1}/S_{\mathbf{c},k}\xrightarrow{D}N(0,1)$.

Let $\eta_{it}=\left[\beta_i^0(t/T)-\pi_i^0B(t/T)\right]'X_{it}$, $\tilde{\eta}_{it}=\eta_{it}-\frac{1}{T}\sum_{t=1}^{T}\eta_{it}$, and $\tilde{u}_{it}=u_{it}-\frac{1}{T}\sum_{t=1}^{T}u_{it}$. Let $\mathbf{c}$ be a nonrandom $p\times 1$ vector with $\|\mathbf{c}\|=1$ and $b_{\mathbf{c}}=\mathbf{c}\otimes B(v)$. Noting that $\tilde{Y}_{it}=\tilde{Z}_{it}'\text{vec}(\pi_i^0)+\tilde{e}_{it}=\tilde{Z}_{it}'\text{vec}(\omega_k^0)+\tilde{Z}_{it}'\text{vec}(\pi_i^0-\omega_k^0)+\tilde{u}_{it}+\tilde{\eta}_{it}$ and $\text{vec}(A_1A_2A_3)=(A_3'\otimes A_1)\text{vec}(A_2)$ (e.g., Bernstein 2005, p.249), we have

$$\begin{aligned}\mathbb{D}_{1k}'\mathbf{c}&=\sqrt{N_kT/J}B(v)'\left(\hat{\omega}_{\hat{G}_k}-\omega_k^0\right)\mathbf{c}=\sqrt{N_kT/J}b_{\mathbf{c}}'\text{vec}\left(\hat{\omega}_{\hat{G}_k}-\omega_k^0\right)\\&=b_{\mathbf{c}}'\left(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{u}_{it}+b_{\mathbf{c}}'\left(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{\eta}_{it}\\&+b_{\mathbf{c}}'\left(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\text{vec}(\pi_i^0-\omega_k^0)\\&\equiv\mathbb{D}_{1k,1}+\mathbb{D}_{1k,2}+\mathbb{D}_{1k,3},\ \text{say.}\end{aligned}$$

It suffices to prove $\mathbb{D}_{1k}/S_{\mathbf{c},k} \xrightarrow{D} N(0,1)$ by showing: (ii1) $\mathbb{D}_{1k,1}/S_{\mathbf{c},k} \xrightarrow{D} N(0,1)$, (ii2) $\mathbb{D}_{1k,2}/S_{\mathbf{c},k} = o_P(1)$, and (ii3) $\mathbb{D}_{1k,3}/S_{\mathbf{c},k} = o_P(1)$.

By arguments as used in the proof of Lemma A.6(i), $\mathbb{D}_{1k,1}/S_{\mathbf{c},k} = \bar{\mathbb{D}}_{1k,1}/S_{\mathbf{c},k} + o_P(1)$, where $\bar{\mathbb{D}}_{1k,1} = b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^T \tilde{Z}_{it}\tilde{u}_{it}$. By Lemma A.8(ii), $\bar{\mathbb{D}}_{1k,1}/S_{\mathbf{c},k} \xrightarrow{D} N(0,1)$. Thus (ii1) follows. To prove (ii2)-(ii3), in view of the fact that $S_{\mathbf{c},k} \asymp \|b_{\mathbf{c}}\|$ by Lemma A.8(i), it suffices to show that $\bar{\mathbb{D}}_{1k,l} = \mathbb{D}_{1k,l}/\|b_{\mathbf{c}}\| = o_P(1)$ for $l = 2,3$. Let $\bar{b}_{\mathbf{c}} = b_{\mathbf{c}}/\|b_{\mathbf{c}}\|$. By arguments as used in the proof of Lemma A.6(i), we can readily show that $\bar{\mathbb{D}}_{1k,2} = \mathbb{D}_{1k,2}^* + o_P(1)$, where $\mathbb{D}_{1k,2}^* = \bar{b}_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^T \tilde{Z}_{it}\tilde{\eta}_{it}$. By Lemma A.9, $\mathbb{D}_{1k,2}^* = o_P(1)$. Thus $\bar{\mathbb{D}}_{1k,2} = o_P(1)$ and (ii2) follows.

By the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \setminus G_k^0\} - \mathbf{1}\{i \in G_k^0 \setminus \hat{G}_k\}$ and that $\pi_i^0 = \omega_k^0$ for $i \in G_k^0$, we have

$$
\begin{aligned}
\bar{\mathbb{D}}_{1k,3} &= \bar{b}_{\mathbf{c}}'\left(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in\hat{G}_k}\sum_{t=1}^T \tilde{Z}_{it}\tilde{Z}_{it}'\mathrm{vec}(\pi_i^0-\omega_k^0) \\
&= \bar{b}_{\mathbf{c}}'\left(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in\hat{G}_k\setminus G_k^0}\sum_{t=1}^T \tilde{Z}_{it}\tilde{Z}_{it}'\mathrm{vec}(\pi_i^0-\omega_k^0).
\end{aligned}
$$

It follows that for any $\epsilon > 0$, we have $P\left(\left|\bar{\mathbb{D}}_{1k,3}\right| \geq \epsilon\right) \leq P(\hat{F}_{kNT}) \to 0$ by Theorem 4.3. That is, $\bar{\mathbb{D}}_{1k,3} = o_P(1)$ and (ii3) follows.

In sum, we have proved that

$$
\sqrt{N_kT/J}\mathbf{c}'\left[\hat{\alpha}_{\hat{G}_k}(v) - \alpha_k^0(v)\right]/S_{\mathbf{c},k} \xrightarrow{D} N(0,1)
$$

where $S_{\mathbf{c},k}^2 = (\mathbf{c}' \otimes B(v)')(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\{\frac{J}{N_kT}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}(\mathbf{c}\otimes B(v)) = \mathbf{c}'\mathbb{S}_k\mathbf{c}$. Then $\sqrt{N_kT/J}\mathbb{S}_k^{-1/2}\left[\hat{\alpha}_{\hat{G}_k}(v) - \alpha_k^0(v)\right] \xrightarrow{D} N(0,\mathbb{I}_p)$ by the Cramér-Wold device. ■

Proof of Theorem 4.5. Let $\mathcal{K} = \{1,2,...,K_{\max}\}$. We partition $\mathcal{K}$ as follows: $\mathcal{K}_0 = \{K \in \mathcal{K} : K = K_0\}$, $\mathcal{K}_- = \{K \in \mathcal{K} : K < K_0\}$, and $\mathcal{K}_+ = \{K \in \mathcal{K} : K > K_0\}$, denoting subsets of $\mathcal{K}$ in which true, under-, and over-fitted models are produced. We prove the theorem by showing that $P\left(\inf_{K\in\mathcal{K}_-\cup\mathcal{K}_+} IC(K,\lambda) > IC(K_0,\lambda)\right) \to 1$ as $(N,T) \to \infty$, i.e., neither the under-fitted model nor the over-fitted model can minimize the information criterion function. Using Theorems 4.3 and 4.4 and Assumption A5, we can readily show that when $K = K_0$

$$
\begin{aligned}
IC(K_0,\lambda) &= \ln\left[\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}\right] + \rho_{NT}JpK_0 \\
&= \ln\left[\frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)}\sum_{t=1}^T\left(\tilde{Y}_{it} - \hat{\omega}_{\hat{G}_k(K_0,\lambda)}'\tilde{Z}_{it}\right)^2\right] + o(1) \xrightarrow{P} \ln\left(\sigma_0^2\right).
\end{aligned}
$$

We consider the cases of under- and over-fitted models separately.

When the model is under-fitted, i.e., $K < K_0$, we have

$$\begin{aligned}
\hat{\sigma}^2_{\hat{G}(K,\lambda)} &= \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\hat{\omega}'_{\hat{G}_k(K,\lambda)}\tilde{Z}_{it}\right)^2 \\
&\geq \min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\hat{\alpha}'_{G_{K,k}}\tilde{Z}_{it}\right)^2 = \min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}}.
\end{aligned}$$

Then by Assumptions A4-A5 and Slutsky Lemma, $\min_{1\leq K<K_0} IC(K,\lambda) \geq \min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K} \ln(\hat{\sigma}^2_{G^{(K)}})+\rho_{NT}JpK \xrightarrow{P} \ln(\underline{\sigma}^2) > \ln(\sigma_0^2)$. It follows that $P\left(\min_{K\in\mathcal{K}_-} IC(K,\lambda) > IC(K_0,\lambda)\right) \to 1$.

Let $\delta_{NT} = (NT/J)^{1/2}$. When the model is over-fitted, by Lemma A.10 and the fact that $\delta^2_{NT}\rho_{NT}J = NT\rho_{NT} \to \infty$ under Assumption A5, we have

$$\begin{aligned}
& P\left(\min_{K\in\mathcal{K}_+} IC(K,\lambda) > IC(K_0,\lambda)\right) \\
= \; & P\left(\min_{K\in\mathcal{K}_+}\left[\delta^2_{NT}\ln\left(\hat{\sigma}^2_{\hat{G}(K,\lambda)}/\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}\right)+\delta^2_{NT}\rho_{NT}Jp(K-K_0)\right]>0\right) \\
= \; & P\left(\min_{K\in\mathcal{K}_+}\delta^2_{NT}\left(\hat{\sigma}^2_{\hat{G}(K,\lambda)}-\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}\right)/\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}+\delta^2_{NT}\rho_{NT}Jp(K-K_0)+o_P(1)>0\right) \\
\to \; & 1 \text{ as } (N,T)\to\infty. \blacksquare
\end{aligned}$$

SUPPLEMENTARY MATERIALS

The online supplementary appendix presents the proofs of the technical lemmas and numerical algorithm in the paper.

ACKNOWLEDGMENTS

The authors thank a co-editor, an associate editor, and three anonymous referees for many constructive comments on the early version of the paper. Su acknowledges financial support from the Singapore Ministry of Education for Academic Research Fund (AcRF) under the Tier 2 Grant MOE2012-T2-2-021 and funding support provided by the Lee Kong Chian Fund for Excellence. Wang acknowledges financial supports from the National Science Foundation of China (No. 71401160, No. 71371160), the Ministry of Education of Humanities and Social Sciences Project of China (No. 14YJC790120), and Fujian Provincial Key Laboratory of Statistics (Xiamen University, No. 2016003).

REFERENCE

Ando, T. and Bai, J. (2016), "Panel Data Models with Grouped Factor Structure under Unknown Group Membership," *Journal of Applied Econometrics*, 31, 163-191.

Andrews, D.W.K. (1984), "Non-strong mixing Autoregressive Processes," *Journal of Applied Probability*, 21, 930-934.

Atak, A., Linton, O. and Xiao, Z. (2011), "A Semiparametric Panel Model for Unbalanced Data with Application to Climate Change in the United Kingdom," *Journal of Econometrics*, 164, 92-115.

Bai, J. (2010), "Common Breaks in Means and Variances for Panel Data," *Journal of Econometrics*, 157, 78-92.

Bai, J. and Lluís Carrion-I-Silvestre, J. (2009), "Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data," *Review of Economic Studies*, 76, 471-501.

Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Barro, R.J. (1991), "Economic Growth in A Cross Section of Countries," *Quarterly Journal of Economics*, 106, 407-443.

Barro, R.J. (1996), *Determinants of Economic Growth: A Cross-country Empirical Study.* The MIT Press, Cambridge.

Bernstein, D. S. (2005), *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Bertsekas, D. (1995), *Nonlinear Programming.* Athena Scientific, Belmont, MA.

Bondell, H. D. and Reich, B. J. (2008), "Simultaneous Regression Shrinkage, Variable Selection, and Supervised Clustering of Predictors with OSCAR," *Biometrics*, 64, 115-123.

Bonhomme, S. and Manresa, E. (2015), "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147-1184.

Cai, Z. (2007), "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, 136, 163-188.

Cai, Z., Chen, L. and Fang, Y. (2014), "Quantile Panel Data Models with Partially Varying Coefficients: Estimating the Growth Effect of FDI," *Working Paper*, Xiamen University.

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941-956.

Cass, D. (1965), "Optimum Growth in An Aggregative Model of Capital Accumulation," *Review of Economic Studies*, 32, 233-240.

Chen, J., Gao, J. and Li, D. (2012), "Semiparametric Trending Panel Data Models with Cross-sectional Dependence," *Journal of Econometrics*, 171, 71-85.

Chen, J., Gao, J. and Li, D. (2013), "Estimation in Partially Linear Single-Index Panel Data Models With Fixed Effects," *Journal of Business & Economic Statistics*, 31, 315-330.

Chen, B. and Huang, L. (2014), "Nonparametric Testing for Smooth Structural Changes in Panel Data Models," *Working Paper*, Department of Economics, University of Rochester.

Chen, X. (2007), "Large Sample Sieve Estimation of Semi-nonparametric Models," In: Heckman, J., and Leamer, E. (Eds.), *Handbook of Econometrics*, Vol. 6B, 5549-5632.

de Boor, C. (2001), *A Practical Guide to Splines.* Springer-Verlag, New York.

DeVore, R. A. and Lorentz, G. G. (1993), *Constructive Approximation.* Springer-Verlag, New York.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2003), *Analysis of Longitudinal Data*, Oxford University Press, Oxford.

Durlauf, S., Kourtellos, A. and Minkin, A., (2001), "The Local Solow Growth Model," *European Economic Review*, 45, 928-940.

Fan, J. and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle properties," *Journal of the American Statistical Association,* 96, 1348-1360.

Fleissig, A.R. and Strauss, J. (1999), "Is OECD Real Per Capita GDP Trend or Difference Stationary? Evidence from Panel Unit Root Tests," *Journal of Macroeconomics*, 21, 673-690.

Hahn, J. and Kuersteiner, G. (2011), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory*, 27, 1152-1191.

Han, A.K. and Park, D. (1989), "Testing for Structural Change in Panel Data: Application to Study of U.S. Foreign Trade in Manufacturing Goods," *Review of Economics and Statistics*, 71, 135-142.

Hsiao, C. (2014), *Analysis of Panel Data.* Cambridge University Press, Cambridge.

Hsiao, C. and Pesaran, M. (2008), "Random Coefficient Models," In: L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, pp. 187-216, 3rd Edition. Springer-Verlag, Berlin.

Huang, J.Z. and Shen, H. (2004), "Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach," *Scandinavian Journal of Statistics*, 31, 515-534.

Huang, J.Z., Wu, C. O., and Zhou, L. (2004), "Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data," *Statistica Sinica*, 14, 763-788.

Juárez, M.A. and Steel, M. F. J. (2010), "Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions," *Journal of Business & Economic Statistics*, 28, 52-66.

Kottaridi, C. and Stengos, T. (2010), "Foreign Direct Investment, Human Capital and Nonlinearities in Economic Growth," *Journal of Macroeconomics*, 32, 858-871.

Ke, Z., Fan, J., and Wu, Y. (2015), "Homogeneity Pursuit," *Journal of American Statistical Association*, 110, 175-194.

Kim, D. (2011), "Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence," *Journal of Econometrics,* 164, 310-330.

Kose, M.A., Prasad, E.S. and Terrones, M.E. (2003), "How Does Globalization Affect the Synchronization of Business Cycles?" *American Economic Review*, 93, 57-62.

Lee, K., Pesaran, M. and Smith, R. (1997), "Growth and Convergence in a Multi-country Empirical Stochastic Solow Model," *Journal of Applied Econometrics*, 12, 357-392.

Li, D., Chen, J. and Gao, J. (2011), "Nonparametric Time-varying Coefficient Panel Data Models with Fixed Effects," *Econometrics Journal*, 14, 387-408.

Li, D., Qian, J. and Su, L. (2016), "Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks," *Journal of the American Statistical Association*, 111, 1804-1819.

Lin, C.-C. and Ng, S. (2012), "Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership Is Unknown ," *Journal of Econometric Methods*, 1, 42-55.

Lluís Carrion-I-Silvestre, J., Barrio-Castro, D., and López-Bazo, E. (2005), "Breaking the Panels: An Application to the GDP Per Capita," *Econometrics Journal*, 8, 159-175.

Moon, H. R. and Weidner, M. (2015), "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543-1579.

Qian, J. and Su, L. (2016), "Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso," *Journal of Econometrics*, 191, 86-109.

Robinson, P.M. (1989), "Nonparametric Estimation of Time-Varying Parameters," In: P. Hackl (eds), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253-264, Springer, Berlin.

Robinson, P.M. (1991), "Time-varying Nonlinear Regression," In: Hackl, P. and Westlund A.H. (Eds.), *Economic Structure Change Analysis and Forecasting,* pp. 179-190, Springer, Berlin.

Robinson, P.M. (2012), "Nonparametric Trending Regression with Cross-sectional Dependence," *Journal of Econometrics*, 169, 4-14.

Schumaker, L. (1981), *Spline Functions: Basic Theory.* Wiley, New York.

Solow, R.M. (1956), "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70, 65-94.

Shen, X., and Huang, H-C. (2010), "Group Pursuit Through a Regularization Solution Surface," *Journal of American Statistical Association*, 105, 727-739.

Su, L. and Chen, Q. (2013), "Testing Homogeneity in Panel Data Models with Interactive Fixed Effects," *Econometric Theory*, 29, 1079-1135.

Su, L. and Jin, S. (2012), "Sieve Estimation of Panel Data Models with Cross Section Dependence," *Journal of Econometrics*, 169, 34-47.

Su, L. and Ju, G. (2017), "Identifying Latent Grouped Effects in Panel Data Models with Interactive Fixed Effects," *Journal of Econometrics*, forthcoming.

Su, L., Shi, Z. and Phillips, P. C. B. (2016), "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215-2264..

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005), "Sparsity and Smoothness via the Fused Lasso," *Journal of the Royal Statistical Society, Series B,* 67, 91-108.

Wang, W., Phillips, P. C. B., and Su, L. (2017), "Homogeneity Pursuit in Panel Data Models: Theory and Applications," *Cowles Foundation Discussion Paper No. 2063*, Yale University.

Xue, L. and Yang, L. (2006), "Additive Coefficient Modeling via Polynomial Spline," *Statistica Sinica* 16, 1423-1446.

Zhang, Y., Su, L., and Phillips, P. C. B. (2011), "Testing for Common Trends in Semi-parametric Panel Data Models with Fixed Effects," *Econometrics Journal*, 15, 56-100.

Zhu, Y., Shen, X., and Pan, W. (2013), "Simultaneous Grouping Pursuit and Feature Selection over an Undirected Graph," *Journal of American Statistical Association*, 108, 713-725.

Supplementary Appendix to
" Sieve Estimation of Time-Varying Panel Data Models with Latent Structures" (NOT for Publication)

Liangjun Su[a], Xia Wang[b], and Sainan Jin[a]
[a] School of Economics, Singapore Management University
[b] Lingnan (University) College, Sun Yat-Sen University

This appendix provides the proofs of some technical lemmas and the numerical algorithm used in the above paper.

B Proofs of the Technical Lemmas

Proof of Lemma A.1. Noting that the eigenvalues of $E\left(X_{it}X_{it}'\right)$ are uniformly bounded away from zero and above from the infinity under Assumption A1(iv), we have

$$\left\|\mathbf{g}\right\|_i^2 = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}\left(t/T\right)' E\left(X_{it}X_{it}'\right)\mathbf{g}\left(t/T\right) \asymp \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}\left(t/T\right)'\mathbf{g}\left(t/T\right).$$

By the property of Riemann integral,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}\left(t/T\right)'\mathbf{g}\left(t/T\right) = \sum_{l=1}^{p}\int_0^1 g_l\left(v\right)^2 dv\left\{1+O\left(\frac{1}{T}\right)\right\} \asymp \sum_{l=1}^{p}\left\|g_l\right\|_2^2.$$

In addition, by Property (iii) of B-splines

$$\sum_{l=1}^{p}\left\|g_l\right\|_2^2 = \sum_{l=1}^{p}\pi_{i,l}'\int_0^1 B\left(v\right)B\left(v\right)' dv\pi_{i,l} \asymp J^{-1}\sum_{l=1}^{p}\left\|\pi_{i,l}\right\|^2 = J^{-1}\left\|\text{vec}\left(\pi_i\right)\right\|^2.$$

This completes the proof of the lemma. ■

Proof of Lemma A.2. The proof of (i) is analogous to that of Lemma A.2 in Huang, Wu, and Zhou (2004) and that of Lemma A.1 in Huang and Shen (2004). Specifically, Huang and Shen (2004) prove (i) for $j=2$ and strictly stationary strong mixing processes without taking the supremum over i. The strict stationarity condition can be relaxed as in Qian and Su (2016). One can readily modify the proofs in these papers to obtain the above claims under the conditions stated in Assumption A1. (ii) follows from (i) with $j=2$. ■

Proof of Lemma A.3. (i) Noting that $Z_{it}=X_{it}\otimes B\left(t/T\right)$ and $\tilde{Z}_{it}=Z_{it}-\frac{1}{T}\sum_{t=1}^{T}Z_{it}$, we have

$$\hat{Q}_{i,\tilde{z}\tilde{z}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}' = \frac{1}{T}\sum_{t=1}^{T}Z_{it}Z_{it}' - \frac{1}{T}\sum_{t=1}^{T}Z_{it}\frac{1}{T}\sum_{t=1}^{T}Z_{it}' \equiv D_{1i}-D_{2i}, \text{ say.}$$

Let $\boldsymbol{\varpi} = \left(\varpi_1', ..., \varpi_p'\right)'$ with each $\varpi_l = (\varpi_{l,1}, ..., \varpi_{l,J})'$ being $J \times 1$ and $\|\boldsymbol{\varpi}\| \leq C < \infty$. Let $g_l\left(v; \varpi_l\right) = \varpi_l' B\left(v\right)$ and $\mathbf{g}_{\boldsymbol{\varpi}}\left(v\right) = \left(g_1\left(v; \varpi_1\right), ..., g_p\left(v; \varpi_p\right)\right)'$. We consider two cases: (1) X_{it} contains a random variable, and (2) $X_{it} = 1$.

Case 1: X_{it} contains a random variable. By Lemmas A.1 and A.2, we have that uniformly in i and $\boldsymbol{\varpi}$

$$\boldsymbol{\varpi}' D_{1i} \boldsymbol{\varpi} = \frac{1}{T} \sum_{t=1}^{T} \left\{\mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' X_{it}\right\}^2 = \|\mathbf{g}_{\boldsymbol{\varpi}}\|_{i,T}^2 = \|\mathbf{g}_{\boldsymbol{\varpi}}\|_i^2 \left\{1 + o_P\left(1\right)\right\}, \quad \|\mathbf{g}_{\boldsymbol{\varpi}}\|_i^2 \asymp J^{-1} \|\boldsymbol{\varpi}\|^2,$$

and the maximum eigenvalue of JD_{1i} and thus $J\hat{Q}_{i,\tilde{z}\tilde{z}}$ is bounded above by a positive number, say $\bar{c}_{zz}$, uniformly in i with probability $1 - o(N^{-1})$.

To examine the minimum eigenvalue of $\hat{Q}_{i,\tilde{z}\tilde{z}}$, we also need to study D_{2i}. By Lemma A.2, $\boldsymbol{\varpi}' D_{2i} \boldsymbol{\varpi} = \{\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' X_{it}\}^2 = \{\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' E\left(X_{it}\right)\}^2 \left\{1 + o_P\left(1\right)\right\}$ uniformly in i and $\boldsymbol{\varpi}$. In addition, one can readily show that $\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' E\left(X_{it}\right) \lesssim J^{-1/2} \|\boldsymbol{\varpi}\|$ uniformly in i and $\boldsymbol{\varpi}$. It follows that

$$\boldsymbol{\varpi}' \left(D_{1i} - D_{2i}\right) \boldsymbol{\varpi} = A_{i,\boldsymbol{\varpi}} + o_P\left(J^{-1}\right),$$

where $A_{i,\boldsymbol{\varpi}} = \frac{1}{T} \sum_{t=1}^{T} E[(\mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' X_{it})^2] - [\frac{1}{T} \sum_{t=1}^{T} E(\mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' X_{it})]^2$ and $o_P\left(J^{-1}\right)$ holds uniformly in i and $\boldsymbol{\varpi}$. By the study of D_{1i}, $A_{i,\boldsymbol{\varpi}} \lesssim J^{-1} \|\boldsymbol{\varpi}\|$ uniformly in i and $\boldsymbol{\varpi}$. We now demonstrate $A_{i,\boldsymbol{\varpi}} \asymp J^{-1} \|\boldsymbol{\varpi}\|$ uniformly in i and $\boldsymbol{\varpi}$ by showing that $A_{i,\boldsymbol{\varpi}} \gtrsim J^{-1} \|\boldsymbol{\varpi}\|$ uniformly in i and $\boldsymbol{\varpi}$. We consider two subcases: (1a) X_{it} does not contain 1, and (1b) X_{it} contains 1.

We first consider subcase (1a) where X_{it} does not contain 1. Define for $r \in [0, 1]$ and $t = 1, ..., T$,

$$\mu_i\left(r\right) = \begin{cases} \mathbf{0}_{p\times 1} & \text{if } r = 0 \\ E\left(X_{it}\right) & \text{if } r \in (\frac{t-1}{T}, \frac{t}{T}] \end{cases} \quad \text{and } \Omega_i\left(r\right) = \begin{cases} \mathbf{0}_{p\times p} & \text{if } r = 0 \\ E\left(X_{it} X_{it}'\right) & \text{if } r \in (\frac{t-1}{T}, \frac{t}{T}] \end{cases}.$$

Let $\bar{\Omega}_i\left(r\right) = \Omega_i\left(r\right) - \mu_i\left(r\right) \mu_i\left(r\right)'$. Observe that $\bar{\Omega}_i\left(t/T\right) = \text{Var}(X_{it})$. By the property of Riemann integral, Jensen inequality, (A.1), Lemma A.1, and Assumption A1(iv), we have

$$\begin{aligned}
A_{i,\boldsymbol{\varpi}} &= \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' \Omega_i\left(t/T\right) \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right) - \left\{\frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_{\boldsymbol{\varpi}}\left(t/T\right)' \mu_i\left(t/T\right)\right\}^2 \\
&= \int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \Omega_i\left(r\right) \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right) dr \left\{1 + O\left(T^{-1}\right)\right\} - \left\{\int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \mu_i\left(r\right) dr\right\}^2 \left\{1 + O\left(T^{-1}\right)\right\} \\
&= \int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \Omega_i\left(r\right) \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right) dr - \left\{\int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \mu_i\left(r\right) dr\right\}^2 + O\left(J^{-1} T^{-1}\right) \\
&\geq \int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \Omega_i\left(r\right) \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right) dr - \int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \mu_i\left(r\right) \mu_i\left(r\right)' \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right) dr + O\left(J^{-1} T^{-1}\right) \\
&= \int_0^1 \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)' \bar{\Omega}_i\left(r\right) \mathbf{g}_{\boldsymbol{\varpi}}\left(r\right) dr + O\left(J^{-1} T^{-1}\right) \\
&\geq \min_{i,t} \mu_{\min}\left(\text{Var}\left(X_{it}\right)\right) \|\mathbf{g}_{\boldsymbol{\varpi}}\|_2^2 \asymp J^{-1} \|\boldsymbol{\varpi}\|^2.
\end{aligned}$$

Consequently, the minimum eigenvalue of $J\hat{Q}_{i,\tilde{z}\tilde{z}}$ is bounded away from below by a positive constant, say $\underline{c}_{zz}$, uniformly in i as $(N,T)\to\infty$. Using Lemma A.2, we can further strengthen the result to $P(\underline{c}_{zz}\leq\min_{1\leq i\leq N}\mu_{\min}(J\hat{Q}_{i,\tilde{z}\tilde{z}}))=1-o\left(N^{-1}\right)$.

Now, we consider subcase (1b) where X_{it} contains 1. Possibly after rearranging the elements in X_{it}, we can write $X_{it}=(1,X_{it}^{(2)\prime})'$ where $X_{it}^{(2)}$ is a $(p-1)\times 1$ vector of stochastic regressors. Accordingly, we write $\mathbf{g}_{\boldsymbol{\varpi}}\left(v\right)=(g_1\left(v;\varpi_1\right),\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(v\right)')'$ with $\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(v\right)=(g_2\left(v;\varpi_2\right),$ $...,g_p\left(v;\varpi_p\right))'$ and $\boldsymbol{\varpi}^{(2)}=\left(\varpi_2',...,\varpi_p'\right)'$. Define for $r\in[0,1]$ and $t=1,...,T$,

$$\mu_i^{(2)}\left(r\right)=\begin{cases}\mathbf{0}_{(p-1)\times 1} & \text{if } r=0\\ E\left(X_{it}^{(2)}\right) & \text{if } r\in(\frac{t-1}{T},\frac{t}{T}]\end{cases}\quad\text{and }\Omega_i^{(2)}\left(r\right)=\begin{cases}\mathbf{0}_{(p-1)\times(p-1)} & \text{if } r=0\\ E\left(X_{it}^{(2)}X_{it}^{(2)\prime}\right) & \text{if } r\in(\frac{t-1}{T},\frac{t}{T}]\end{cases}.$$

Let $\bar{\Omega}_i^{(2)}\left(r\right)=\Omega_i^{(2)}\left(r\right)-\mu_i^{(2)}\left(r\right)\mu_i^{(2)}\left(r\right)'$. Noting that for any $r\in(0,1]$,

$$\mu_i\left(r\right)=\begin{pmatrix}1\\ \mu_i^{(2)}\left(r\right)\end{pmatrix},\ \Omega_i\left(r\right)=\begin{pmatrix}1 & \mu_i^{(2)}\left(r\right)\\ \mu_i^{(2)}\left(r\right)' & \Omega_i^{(2)}\left(r\right)\end{pmatrix},\ \bar{\Omega}_i\left(r\right)=\begin{pmatrix}0 & \mathbf{0}_{1\times(p-1)}\\ \mathbf{0}_{(p-1)\times 1} & \bar{\Omega}_i^{(2)}\left(r\right)\end{pmatrix},$$

and $\mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)'\mu_i\left(r\right)=g_1\left(r;\varpi_1\right)+\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)$, we have $A_{i,\boldsymbol{\varpi}}=\bar{A}_{i,\boldsymbol{\varpi}}+O\left(J^{-1}T^{-1}\right)$ with

$$\begin{aligned}\bar{A}_{i,\boldsymbol{\varpi}} &= \int_0^1\mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)'\Omega_i\left(r\right)\mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)dr-\left\{\int_0^1\mathbf{g}_{\boldsymbol{\varpi}}\left(r\right)'\mu_i\left(r\right)dr\right\}^2\\ &= \int_0^1 g_1\left(r;\varpi_1\right)^2dr-\left(\int_0^1 g_1\left(r;\varpi_1\right)dr\right)^2\\ &\quad+\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\Omega_i^{(2)}\left(r\right)\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)dr-\left(\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)dr\right)^2\\ &\quad+2\left\{\int_0^1 g_1\left(r;\varpi_1\right)\mu_i^{(2)}\left(r\right)'\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)dr-\int_0^1 g_1\left(r;\varpi_1\right)dr\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)dr\right\}\\ &= \bar{A}_{i,\boldsymbol{\varpi}}^{(1)}+\bar{A}_{i,\boldsymbol{\varpi}}^{(2)},\end{aligned}$$

where

$$\begin{aligned}\bar{A}_{i,\boldsymbol{\varpi}}^{(1)} &= \int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\bar{\Omega}_i^{(2)}\left(r\right)\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)dr,\\ \bar{A}_{i,\boldsymbol{\varpi}}^{(2)} &= \left[\int_0^1 g_1\left(r;\varpi_1\right)^2dr-\left(\int_0^1 g_1\left(r;\varpi_1\right)dr\right)^2\right]\\ &\quad+\left[\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)\mu_i^{(2)}\left(r\right)\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'dr-\left(\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)dr\right)^2\right]\\ &\quad+2\left\{\int_0^1 g_1\left(r;\varpi_1\right)\mu_i^{(2)}\left(r\right)'\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)dr-\int_0^1 g_1\left(r;\varpi_1\right)dr\int_0^1\mathbf{g}_{\boldsymbol{\varpi}^{(2)}}^{(2)}\left(r\right)'\mu_i^{(2)}\left(r\right)dr\right\}.\end{aligned}$$

Noting that $\mathbf{g}^{(2)}_{\boldsymbol{\varpi}^{(2)}}(r) = (g_2(r;\varpi_2), ..., g_p(r;\varpi_p))' = W^{(2)\prime}B(r)$ where $W^{(2)} = (\varpi_2, ..., \varpi_p)$ is a $J \times (p-1)$ matrix such that $\boldsymbol{\varpi}^{(2)}$ =vec$\big(W^{(2)}\big)$ and tr$(A_1A_2A_3A_4)$ =vec$(A_1)'(A_2 \otimes A_4')$vec(A_3') (e.g., Bernstein 2005, p. 253), we have

$$\begin{aligned}
\bar{A}^{(1)}_{i,\boldsymbol{\varpi}} &= \int_0^1 \mathbf{g}^{(2)}_{\boldsymbol{\varpi}^{(2)}}(r)' \bar{\Omega}^{(2)}_i(r) \mathbf{g}^{(2)}_{\boldsymbol{\varpi}^{(2)}}(r)\,dr = \int_0^1 B(v)' W^{(2)}\bar{\Omega}^{(2)}_i(r) W^{(2)\prime}B(v)\,dr \\
&= \int_0^1 \mathrm{tr}\left(W^{(2)}\bar{\Omega}^{(2)}_i(r) W^{(2)\prime}B(r) B(r)'\right) dr \\
&= \boldsymbol{\varpi}^{(2)\prime}\int_0^1 \bar{\Omega}^{(2)}_i(r) \otimes (B(r) B(r)')dr\boldsymbol{\varpi}^{(2)} \\
&= \boldsymbol{\varpi}'\begin{pmatrix} \mathbf{0}_{J\times J} & \mathbf{0}_{J\times J(p-1)} \\ \mathbf{0}_{J(p-1)\times J} & \int_0^1 \bar{\Omega}^{(2)}_i(r) \otimes (B(r) B(r)')dr \end{pmatrix}\boldsymbol{\varpi}.
\end{aligned}$$

Let $\bar{g}_1(r;\varpi_1) = g_1(r;\varpi_1) - \int_0^1 g_1(v;\varpi_1)\,dv$, $\varphi^{(2)}_i(r) = \mu^{(2)}_i(r) \otimes B(r)$, and $\bar{\varphi}^{(2)}_i(r) = \varphi^{(2)}_i(r) - \int_0^1 \varphi^{(2)}_i(v)\,dv$. Using $\bar{g}_1(r;\varpi_1) = g_1(r;\varpi_1) - \int_0^1 g_1(v;\varpi_1)\,dv = \varpi_1'\bar{B}(r)$ with $\bar{B}(r) = B(r) - \int_0^1 B(v)\,dv$ and $\mu^{(2)}_i(r)'\mathbf{g}^{(2)}_{\boldsymbol{\varpi}^{(2)}}(r) = \boldsymbol{\varpi}^{(2)\prime}(\mu^{(2)}_i(r) \otimes B(r)) = \boldsymbol{\varpi}^{(2)\prime}\varphi^{(2)}_i(r)$, we have

$$\begin{aligned}
\bar{A}^{(2)}_{i,\boldsymbol{\varpi}} &= \varpi_1'\int_0^1 \bar{B}(r)\bar{B}(r)'\,dr\varpi_1 + \boldsymbol{\varpi}^{(2)\prime}\int_0^1 \bar{\varphi}^{(2)}_i(r)\bar{\varphi}^{(2)}_i(r)'\,dr\boldsymbol{\varpi}^{(2)} + 2\varpi_1'\int_0^1 \bar{B}(r)\bar{\varphi}^{(2)}_i(r)'\,dr\boldsymbol{\varpi}^{(2)} \\
&= \boldsymbol{\varpi}'\begin{pmatrix} \int_0^1 \bar{B}(r)\bar{B}(r)'\,dr & \int_0^1 \bar{B}(r)\bar{\varphi}^{(2)}_i(r)'\,dr \\ \int_0^1 \bar{\varphi}^{(2)}_i(r)\bar{B}(r)'\,dr & \int_0^1 \bar{\varphi}^{(2)}_i(r)\bar{\varphi}^{(2)}_i(r)'\,dr \end{pmatrix}\boldsymbol{\varpi}.
\end{aligned}$$

It follows that $\bar{A}_{i,\boldsymbol{\varpi}} = \boldsymbol{\varpi}'\bar{A}_i(1)\boldsymbol{\varpi} + \boldsymbol{\varpi}'\bar{A}_i(2)\boldsymbol{\varpi}$, where

$$\begin{aligned}
\bar{A}_i(1) &= \begin{pmatrix} \int_0^1 \bar{B}(r)\bar{B}(r)'\,dr & \int_0^1 \bar{B}(r)\bar{\varphi}^{(2)}_i(r)'\,dr \\ \mathbf{0}_{J(p-1)\times J} & \int_0^1 \bar{\Omega}^{(2)}_i(r) \otimes (B(r) B(r)')dr \end{pmatrix} \text{ and} \\
\bar{A}_i(2) &= \begin{pmatrix} \mathbf{0}_{J\times J} & \mathbf{0}_{J\times J(p-1)} \\ \int_0^1 \bar{\varphi}^{(2)}_i(r)\bar{B}(r)'\,dr & \int_0^1 \bar{\varphi}^{(2)}_i(r)\bar{\varphi}^{(2)}_i(r)'\,dr \end{pmatrix}.
\end{aligned}$$

By Lemma 21.2.1 in Harville (1997), the eigenvalues of either the lower or upper block triangular matrices are given by the collection of the eigenvalues of their diagonal blocks. As a result,

$$\begin{aligned}
\mu_{\min}\left(\bar{A}_i(1)\right) &= \min\left(\mu_{\min}\left(\int_0^1 \bar{B}(r)\bar{B}(r)'\,dr\right), \mu_{\min}\left(\int_0^1 \bar{\Omega}^{(2)}_i(r) \otimes (B(r) B(r)')dr\right)\right), \\
\mu_{\min}\left(\bar{A}_i(2)\right) &= \min\left(\mu_{\min}(\mathbf{0}_{J\times J}), \mu_{\min}\left(\int \bar{\varphi}^{(2)}_i(r)\bar{\varphi}^{(2)}_i(r)'\,dr\right)\right) = 0.
\end{aligned}$$

Noting that $\bar{B}$, as a centered version of B, also shares Property (iii) of B-splines, we have $\mu_{\min}\left(\int_0^1 \bar{B}(r)\bar{B}(r)' dr\right) \gtrsim J^{-1}$. In addition,

$$\begin{aligned}\mu_{\min}\left(\int_0^1 \bar{\Omega}_i^{(2)}(r)\otimes(B(r)B(r)')dr\right) &\geq \min_{i,t}\bar{\Omega}_i^{(2)}(t/T)\,\mu_{\min}\left(\mathbb{I}_{p-1}\otimes\int_0^1 B(r)B(r)'dr\right)\\ &= \min_{i,t}\bar{\Omega}_i^{(2)}(t/T)\,\mu_{\min}\left(\int_0^1 B(r)B(r)'dr\right)\gtrsim J^{-1}.\end{aligned}$$

Consequently, $\mu_{\min}\left(\bar{A}_i(1)\right)\gtrsim J^{-1}$. By Weyl inequality (e.g., Bernstein 2005, Theorem 8.4.11, p.274), $\mu_{\min}\left(\bar{A}_i(1)+\bar{A}_i(2)\right)\geq \mu_{\min}\left(\bar{A}_i(1)\right)+\mu_{\min}\left(\bar{A}_i(2)\right)\gtrsim J^{-1}$ and

$$\bar{A}_{i,\boldsymbol{\varpi}}\gtrsim J^{-1}\|\boldsymbol{\varpi}\|.$$

Consequently, the minimum eigenvalue of $J\hat{Q}_{i,\tilde{z}\tilde{z}}$ is bounded away from below by a positive constant, say $\underline{c}_{zz}$, uniformly in i as $(N,T)\to\infty$. Using Lemma A.2, we can further strengthen the result to $P(\underline{c}_{zz}\leq \min_{1\leq i\leq N}\mu_{\min}(J\hat{Q}_{i,\tilde{z}\tilde{z}}))=1-o(N^{-1})$.

Case 2: $X_{it}=1$. In this case, we only need to apply the basic properties of B-splines. In particular,

$$\begin{aligned}\boldsymbol{\varpi}'(D_{1i}-D_{2i})\boldsymbol{\varpi} &= \boldsymbol{\varpi}'\left[\frac{1}{T}\sum_{t=1}^T B(t/T)B(t/T)'-\frac{1}{T}\sum_{t=1}^T B(t/T)\frac{1}{T}\sum_{t=1}^T B(t/T)'\right]\boldsymbol{\varpi}\\ &\asymp \boldsymbol{\varpi}'\left[\int_0^1 B(v)B(v)'dv-\int_0^1 B(v)dv\int_0^1 B(v)'dv\right]\boldsymbol{\varpi}\\ &= \boldsymbol{\varpi}'\int_0^1 \bar{B}(v)\bar{B}(v)'dv\boldsymbol{\varpi},\end{aligned}$$

where $\bar{B}(v)=B(v)-\int_0^1 B(s)ds$. By the properties of B-splines, $J\int_0^1 \bar{B}(v)\bar{B}(v)'dv$ has maximum eigenvalue bounded above by M_2 given in (A.1) and minimum eigenvalue bounded below from zero. Then the conclusion in (i) also holds.

(ii) By (i) and the fact that for a $k\times k$ symmetric matrix A, $\underline{c}\mathbb{I}_k\leq A\leq\bar{c}\mathbb{I}_k$ if and only if $\underline{c}\leq\mu_{\min}(A)\leq\mu_{\max}(A)\leq\bar{c}$ (e.g., Bernstein 2005, Lemma 8.4.1., p.271), we have $P(\underline{c}_{zz}\mathbb{I}_{Jp}\leq J\hat{Q}_{i,\tilde{z}\tilde{z}}\leq\bar{c}_{zz}\mathbb{I}_{Jp})=1-o(N^{-1})$. By the fact that for symmetric matrices $A=\{a_{ij}\}$ and $B=\{b_{ij}\}$, $A\leq B$ implies that $a_{ii}\leq b_{ii}$ for each i (see, e.g., Bernstein 2005, Fact 8.8.9, p. 296). Consequently, $P(\underline{c}_{zz}\mathbb{I}_{Jp}\leq\mathrm{diag}(J\hat{Q}_{i,\tilde{z}\tilde{z}})\leq\bar{c}_{zz}\mathbb{I}_{Jp})=1-o(N^{-1})$ and (ii) follows. ■

Proof of Lemma A.4. (i) Recall that $a_{it}=\beta_i^0(t/T)-\pi_i^{0\prime}B(t/T)$, $\eta_{it}=X_{it}'a_{it}$, and $\vartheta_{NT}=\max_{1\leq k\leq K}\sup_{v\in[0,1]}\|\alpha_k^0(v)-\omega_k^{0\prime}B(v)\|$. Noting that $e_{it}=u_{it}+X_{it}'\left[\beta_i^0(t/T)-\pi_i^{0\prime}B(t/T)\right]=u_{it}+\eta_{it}$, we have $\hat{Q}_{i,\tilde{z}\tilde{e}}=\hat{Q}_{i,\tilde{z}\tilde{\eta}}+\hat{Q}_{i,\tilde{z}\tilde{u}}$, where $\hat{Q}_{i,\tilde{z}\tilde{u}}=\frac{1}{T}\sum_{t=1}^T\tilde{Z}_{it}\tilde{u}_{it}$, $\hat{Q}_{i,\tilde{z}\tilde{\eta}}=\frac{1}{T}\sum_{t=1}^T\tilde{Z}_{it}\tilde{\eta}_{it}$, and $\tilde{u}_{it}$ and $\tilde{\eta}_{it}$ are defined analogously to $\tilde{Z}_{it}$. Note that $\tilde{e}_{it}=\tilde{u}_{it}+\tilde{\eta}_{it}$. By the property of

B-splines,

$$
\begin{aligned}
E\left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}\eta_{it}\right\|^{2} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[\left(X_{it}\otimes B\left(t/T\right)\right)'\left(X_{is}\otimes B\left(s/T\right)\right)\eta_{it}\eta_{is}\right] \\
&\leq \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(X_{it}'X_{is}X_{it}'a_{it}X_{is}'a_{is}\right)B\left(t/T\right)'B\left(s/T\right) \\
&\leq \bar{c}_{x}^{4/q}\vartheta_{NT}^{2}\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}B\left(t/T\right)'B\left(s/T\right) \\
&\leq \bar{c}_{x}^{4/q}\vartheta_{NT}^{2}\sum_{j=1}^{J}\left\{\frac{1}{T}\sum_{t=1}^{T}B_{j}\left(t/T\right)\frac{1}{T}\sum_{s=1}^{T}B_{j}\left(s/T\right)\right\} \\
&\asymp \bar{c}_{x}^{4/q}\vartheta_{NT}^{2}\sum_{j=1}^{J}\left\{\int B_{j}\left(v\right)dv\right\}^{2}=O\left(J^{-2\gamma}\right)O\left(J^{-1}\right)=O\left(J^{-2\gamma-1}\right),
\end{aligned}
$$

where we use the fact that $E\left(X_{it}'X_{is}X_{it}'a_{it}X_{is}'a_{is}\right)\leq\vartheta_{NT}^{2}E\left(\left\|X_{it}\right\|^{2}\left\|X_{is}\right\|^{2}\right)\leq\vartheta_{NT}^{2}E(\left\|X_{it}\right\|^{4})\leq\bar{c}_{x}^{4/q}\vartheta_{NT}^{2}=O\left(J^{-2\gamma}\right)$ by Assumptions A1(iii) and (v) and (4.1). It follows that $\left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}\eta_{t}\right\|=O_{P}\left(J^{-\gamma-1/2}\right)$. Similarly, we can show that $\left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}\right\|=O_{P}\left(J^{-1/2}\right)$ and $\frac{1}{T}\sum_{t=1}^{T}\eta_{it}=O_{P}\left(J^{-\gamma}\right)$. Then $\left\|\hat{Q}_{i,\tilde{z}\tilde{\eta}}\right\|=\left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}\eta_{it}-\frac{1}{T}\sum_{t=1}^{T}Z_{it}\frac{1}{T}\sum_{t=1}^{T}\eta_{it}\right\|=O\left(J^{-\gamma-1/2}\right)$.

Noting that

$$
\begin{aligned}
E\left\|\frac{1}{T}\sum_{t=1}^{T}\left(X_{it}u_{it}\right)\otimes B\left(t/T\right)\right\|^{2} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(X_{it}'X_{is}u_{it}u_{is}\right)B\left(t/T\right)'B\left(s/T\right) \\
&= \frac{1}{T^{2}}\sum_{t=1}^{T}E\left(X_{it}'X_{it}u_{it}^{2}\right)\left\|B\left(t/T\right)\right\|^{2} \\
&\quad +\frac{2}{T^{2}}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E\left(X_{it}'X_{is}u_{it}u_{is}\right)B\left(t/T\right)'B\left(s/T\right) \\
&\equiv A_{i1}+A_{i2},\ \text{say.}
\end{aligned}
$$

By Assumptions A1(iii)-(iv), Jensen inequality, and the properties of B-splines,

$$
A_{i1}\leq\frac{\left(\bar{c}_{x}\bar{c}_{u}\right)^{q/4}}{T^{2}}\sum_{t=1}^{T}\left\|B\left(t/T\right)\right\|^{2}\asymp\frac{\left(\bar{c}_{x}\bar{c}_{u}\right)^{q/4}}{T}\int\left\|B\left(v\right)\right\|^{2}dv=O\left(T^{-1}\right).
$$

Noting that $\left\|B\left(v\right)\right\|=\{\sum_{j=-d+1}^{J_{0}}B_{j}\left(v\right)^{2}\}^{1/2}\leq\{\sum_{j=-d+1}^{J_{0}}B_{j}\left(v\right)\}^{1/2}=1$ uniformly in $v\in$

$[0,1]$, by Assumptions A1(ii)-(iv) and Davydov inequality for strong mixing processes,

$$\begin{aligned}
|A_{i2}| &\leq \frac{2}{T^2}\sum_{j=1}^{p}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\left|E\left(X_{it,j}u_{it}X_{is,j}u_{is}\right)B\left(t/T\right)'B\left(s/T\right)\right| \\
&\leq \frac{16}{T^2}\sum_{j=1}^{p}\max_{i,t}\left\{E\left|X_{it,j}'u_{it}\right|^{q/2}\right\}^{4/q}\sum_{t=1}^{T-1}\sum_{l=1}^{\infty}\alpha\left(l\right)^{(q-4)/q}=O\left(T^{-1}\right).
\end{aligned}$$

Hence $E\left\|\frac{1}{T}\sum_{t=1}^{T}\left(X_{it}u_{it}\right)\otimes B\left(t/T\right)\right\|^2=O\left(T^{-1}\right)$ and $\left\|\frac{1}{T}\sum_{t=1}^{T}\left(X_{it}u_{it}\right)\otimes B\left(t/T\right)\right\|=O_P\left(T^{-1/2}\right)$ by Chebyshev inequality. It follows that

$$\begin{aligned}
\left\|\hat{Q}_{i,\tilde{z}\tilde{u}}\right\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}u_{it}-\frac{1}{T}\sum_{t=1}^{T}Z_{it}\frac{1}{T}\sum_{t=1}^{T}u_{it}\right\| \\
&\leq \left\|\frac{1}{T}\sum_{t=1}^{T}\left(X_{it}u_{it}\right)\otimes B\left(t/T\right)\right\|+\left\|\frac{1}{T}\sum_{t=1}^{T}Z_{it}\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}u_{it}\right\| \\
&= O_P\left(T^{-1/2}\right)+O_P\left(J^{-1/2}\right)O_P\left(T^{-1/2}\right)=O_P\left(T^{-1/2}\right).
\end{aligned}$$

Consequently, $||\hat{Q}_{i,\tilde{z}\tilde{e}}|| \leq ||\hat{Q}_{i,\tilde{z}\tilde{\eta}}||+||\hat{Q}_{i,\tilde{z}\tilde{u}}|| = O_P\left(J^{-\gamma-1/2}+T^{-1/2}\right)$. This completes the proof of (i).

(ii) The proof of (ii) is analogous to that of (i) and thus omitted.

(iii) Using the fact that $\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{\eta}_{it}=\frac{1}{T}\sum_{t=1}^{T}Z_{it}\tilde{\eta}_{it}$ and by the triangle and Jensen inequalities,

$$\left\|\hat{Q}_{i,\tilde{z}\tilde{\eta}}\right\|^2\leq\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|Z_{it}\right\|\left|\tilde{\eta}_{it}\right|\right\}^2\leq\frac{1}{T}\sum_{t=1}^{T}\left\|Z_{it}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}_{it}^2.$$

By (4.1)-(4.2),

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}_{it}^2\leq\frac{1}{T}\sum_{t=1}^{T}\eta_{it}^2=\frac{1}{T}\sum_{t=1}^{T}X_{it}'X_{it}\left[\beta_i^0\left(t/T\right)-\pi_i^{0\prime}B(t/T)\right]^2\leq\vartheta_{NT}^2\frac{1}{T}\sum_{t=1}^{T}\left\|X_{it}\right\|^2.$$

Using the same arguments as used in the study of $A_{i3,1}$ in the proof of Lemma A.5 below, we can show that $P(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\left(\left\|X_{it}\right\|^2-E\left\|X_{it}\right\|^2\right)\geq\epsilon)=o\left(N^{-1}\right)$. It follows that

$$\begin{aligned}
P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\left\|X_{it}\right\|^2\geq 2\bar{c}_{xx}\right) &\leq P\left(\max_{i,t}E\left\|X_{it}\right\|^2+\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\left(\left\|X_{it}\right\|^2-E\left\|X_{it}\right\|^2\right)\geq 2\bar{c}_{xx}\right) \\
&= o\left(N^{-1}\right)
\end{aligned}$$

where $\bar{c}_{xx}\equiv\max_{i,t}E\left\|X_{it}\right\|^2\leq\left(\bar{c}_x\right)^{2/q}$ under Assumption A1(iii). Then

$$P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}_{it}^2\geq 2\bar{c}_{xx}\vartheta_{NT}^2\right)=P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\left\|X_{it}\right\|^2\geq 2\bar{c}_{xx}\right)=o\left(N^{-1}\right). \quad \text{(B.1)}$$

By fact that $\|B(v)\| \leq 1$ uniformly in $v \in [0,1]$, we have $\|Z_{it}\| = \|X_{it} \otimes B(t/T)\| = \|X_{it}\| \|B(t/T)\| \leq \|X_{it}\|$. It follows that

$$P\left(\max_{1\leq i\leq N} \frac{1}{T}\sum_{t=1}^{T} \|Z_{it}\|^2 \geq 2\bar{c}_{xx}\right) \leq P\left(\max_{1\leq i\leq N} \frac{1}{T}\sum_{t=1}^{T} \|X_{it}\|^2 \geq 2\bar{c}_{xx}\right) = o\left(N^{-1}\right).$$

Consequently, we have

$$P\left(\max_{1\leq i\leq N} \left\|\hat{Q}_{i,\tilde{z}\tilde{\eta}}\right\| \geq 2\left(\bar{c}_x\right)^{2/q} \vartheta_{NT}\right) = o\left(N^{-1}\right). \tag{B.2}$$

(iv) Now, we study the uniform probability order of $\left\|\hat{Q}_{i,\tilde{z}\tilde{u}}\right\|$. Note that $\hat{Q}_{i,\tilde{z}\tilde{u}} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}u_{it}$ $-\frac{1}{T}\sum_{t=1}^{T} Z_{it}\frac{1}{T}\sum_{t=1}^{T} u_{it}$. Following analogous analysis of $A_{i3,1}$ in the proof of Lemma A.5 below, we can show that for any $\epsilon > 0$,

$$\begin{aligned}
P\left(\max_{1\leq i\leq N} \left\|\frac{1}{T}\sum_{t=1}^{T} Z_{it}u_{it}\right\| \geq \epsilon T^{-1/2}(\ln T)^3\right) &= o\left(N^{-1}\right), \\
P\left(\max_{1\leq i\leq N} \left\|\frac{1}{T}\sum_{t=1}^{T} u_{it}\right\| \geq \epsilon T^{-1/2}(\ln T)^3\right) &= o\left(N^{-1}\right), \\
P\left(\max_{1\leq i\leq N} \left\|\frac{1}{T}\sum_{t=1}^{T} [Z_{it} - E(Z_{it})]\right\| \geq \epsilon T^{-1/2}(\ln T)^3\right) &= o\left(N^{-1}\right).
\end{aligned}$$

These results, in conjunction with the fact that $\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T} E(Z_{it})\right\| = O\left(J^{-1/2}\right)$, imply that $P\left(\max_{1\leq i\leq N}\left\|\hat{Q}_{i,\tilde{z}\tilde{u}}\right\| \geq \epsilon T^{-1/2}(\ln T)^3\right) = o\left(N^{-1}\right)$. ■

Proof of Lemma A.5. (i) Using $\tilde{Y}_{it} = \tilde{Z}_{it}'\text{vec}(\pi_i^0) + \tilde{e}_{it}$, we have $\text{vec}(\tilde{\pi}_i - \pi_i^0) = \hat{Q}_{i,\tilde{z}\tilde{z}}^{-1}\hat{Q}_{i,\tilde{z}\tilde{e}}$. By the fact $\|A\| = ||\text{vec}(A)||$ and $\|Ab\| = \|Ab\|_{\text{sp}} \leq \|A\|_{\text{sp}}\|b\|$ for any matrix A and conformable vector b, we have by Lemmas A.3(i) and A.4(iii)-(iv)

$$\begin{aligned}
\max_{1\leq i\leq N}\left\|\tilde{\pi}_i - \pi_i^0\right\| &= \max_{1\leq i\leq N}\left\|\hat{Q}_{i,\tilde{z}\tilde{z}}^{-1}\hat{Q}_{i,\tilde{z}\tilde{e}}\right\| \leq J^{1/2}\max_{1\leq i\leq N}\left\|\left(J\hat{Q}_{i,\tilde{z}\tilde{z}}\right)^{-1}\right\|_{\text{sp}}\max_{1\leq i\leq N}\left\|J^{1/2}\hat{Q}_{i,\tilde{z}\tilde{e}}\right\| \\
&= J^{1/2}O_P(1)\,o_P(1) = o_P\left(J^{1/2}\right).
\end{aligned}$$

(ii) Recall that $\tilde{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}[\tilde{Y}_{it} - \tilde{Z}_{it}'\text{vec}(\tilde{\pi}_i)]^2$ and $\bar{\sigma}_{i,T}^2 = T^{-1}\sum_{t=1}^{T} E(u_{it}^2)$. Noting that $\tilde{Y}_{it} = \tilde{Z}_{it}'\text{vec}(\pi_i^0) + \tilde{e}_{it}$, we have

$$\begin{aligned}
\tilde{\sigma}_i^2 &= \frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it} + \tilde{Z}_{it}'\text{vec}\left(\pi_i^0 - \tilde{\pi}_i\right)]^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^2 + \text{vec}\left(\pi_i^0 - \tilde{\pi}_i\right)'\frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\text{vec}\left(\pi_i^0 - \tilde{\pi}_i\right) + 2\text{vec}\left(\pi_i^0 - \tilde{\pi}_i\right)'\frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_{it}'\tilde{e}_{it} \\
&\equiv A_{i3} + A_{i4} + A_{i5}, \text{ say.}
\end{aligned}$$

We prove the lemma by showing that (ii1) $P(\max_{1\leq i\leq N}|A_{i3}-\bar{\sigma}_{i,T}^2|\geq\epsilon)=o(N^{-1})$, (ii2) $P(\max_{1\leq i\leq N}|A_{i4}|\geq\epsilon)=o(N^{-1})$, and (ii3) $P(\max_{1\leq i\leq N}|A_{i5}|\geq\epsilon)=o(N^{-1})$ for any $\epsilon>0$.

Using $\tilde{e}_{it}=\tilde{u}_{it}+\tilde{\eta}_{it}$ as used in the proof of Lemma A.5, $\tilde{u}_{it}=u_{it}-\frac{1}{T}\sum_{t=1}^T u_{it}$, and the fact that $\sum_{t=1}^T\tilde{u}_{it}\tilde{\eta}_{it}=\sum_{t=1}^T u_{it}\tilde{\eta}_{it}$, we have

$$\begin{aligned}
A_{i3}-\bar{\sigma}_{i,T}^2 &= \left(\frac{1}{T}\sum_{t=1}^T\tilde{u}_{it}^2-\bar{\sigma}_{i,T}^2\right)+\frac{1}{T}\sum_{t=1}^T\tilde{\eta}_{it}^2+\frac{2}{T}\sum_{t=1}^T\tilde{u}_{it}\tilde{\eta}_{it}\\
&= \frac{1}{T}\sum_{t=1}^T\left[u_{it}^2-E\left(u_{it}^2\right)\right]-\left(\frac{1}{T}\sum_{t=1}^T u_{it}\right)^2+\frac{1}{T}\sum_{t=1}^T\tilde{\eta}_{it}^2+\frac{2}{T}\sum_{t=1}^T u_{it}\tilde{\eta}_{it}\\
&\equiv A_{i3,1}-A_{i3,2}+A_{i3,3}+2A_{i3,4},\text{ say.}
\end{aligned}$$

Let $c_{NT}=T(\ln N)^{-\epsilon_0}$ for some $\epsilon_0>1$. Define $\varsigma_{it}^{(1)}=u_{it}^2\mathbf{1}_{it}-E\left(u_{it}^2\mathbf{1}_{it}\right)$, $\varsigma_{it}^{(2)}=u_{it}^2\bar{\mathbf{1}}_{it}$, and $\varsigma_{it}^{(3)}=E\left(u_{it}^2\bar{\mathbf{1}}_{it}\right)$, where $\mathbf{1}_{it}=\mathbf{1}\left\{u_{it}^2\leq c_{NT}\right\}$ and $\bar{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. Then $A_{i3,1}=\frac{1}{T}\sum_{t=1}^T(\varsigma_{it}^{(1)}+\varsigma_{it}^{(2)}-\varsigma_{it}^{(3)})$. Let $\upsilon_0^2=\max_{i,t}[\mathrm{Var}(\varsigma_{it}^{(1)})+2\sum_{s=t+1}^T|\mathrm{Cov}(\varsigma_{it}^{(1)},\varsigma_{is}^{(1)})|]$. Note that $\upsilon_0^2<\infty$ under our mixing and moment conditions in Assumption A1. By Bernstein inequality for strong mixing processes (e.g., Theorem 2 in Merlevéde, Peligrad, and Rio 2009 or Lemma A.2 in Qian and Su 2016), there exists a positive constant C_0 such that for any $\epsilon>0$,

$$\begin{aligned}
P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^T\varsigma_{it}^{(1)}\right|\geq\epsilon\right) &\leq \sum_{i=1}^N P\left(\left|\sum_{t=1}^T\varsigma_{it}^{(1)}\right|\geq T\epsilon\right)\\
&\leq N\exp\left(-\frac{C_0T^2\epsilon^2}{T\upsilon_0^2+c_{NT}^2+\epsilon c_{NT}(\ln T)^2}\right)=o\left(N^{-1}\right).
\end{aligned}$$

By Markov inequality, Lebesgue dominated convergence theorem, and Assumptions A1(iii) and A2(i)

$$\begin{aligned}
P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^T\varsigma_{it}^{(2)}\right|\geq\epsilon\right) &\leq P\left(\max_{1\leq i\leq N}\max_{1\leq t\leq T}u_{it}^2>c_{NT}\right)\\
&\leq \frac{1}{c_{NT}^{q/2}}\sum_{i=1}^N\sum_{t=1}^T E\left[|u_{it}|^q\,\mathbf{1}\left\{u_{it}^2>c_{NT}\right\}\right]=o\left(N^{-1}\right),
\end{aligned}$$

where we use the fact that $N^2Tc_{NT}^{-q/2}=N^2T^{1-q/2}(\ln N)^{q\epsilon_0/2}=o(1)$ under Assumption A2(i). In addition, $\frac{1}{T}\sum_{t=1}^T\varsigma_{it}^{(3)})\leq\max_{1\leq i\leq N}\max_{1\leq t\leq T}E\left(u_{it}^2\bar{\mathbf{1}}_{it}\right)=o(1)$. Consequently $P\left(\max_{1\leq i\leq N}|A_{i3,1}|\geq\epsilon\right)=o\left(N^{-1}\right)$ for any $\epsilon>0$. Analogously, we can show that $P(\max_{1\leq i\leq N}|A_{i3,2}|\geq\epsilon)=o\left(N^{-1}\right)$ for any $\epsilon>0$. By (B.1),

$$P\left(\max_{1\leq i\leq N}|A_{i3,3}|\geq\epsilon\right)=P\left(\vartheta_{NT}^2\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^T\|X_{it}\|^2\geq\epsilon\right)=o\left(N^{-1}\right).$$

Similarly, we can show that

$$P\left(\max_{1\leq i\leq N}|A_{i3,4}|\geq\epsilon\right)=P\left(\max_{1\leq i\leq N}\left(\frac{1}{T}\sum_{t=1}^T u_{it}^2\right)^{1/2}\left\{\frac{1}{T}\sum_{t=1}^T\tilde{\eta}_{it}^2\right\}^{1/2}\geq\epsilon\right)=o(1).$$

In sum, we have $P(\max_{1\leq i\leq N}|A_{i3}-\bar{\sigma}_{i,T}^{2}|\geq\epsilon)=o(N^{-1})$. This shows (ii1).

For (ii2)-(ii3), we observe that

$$\begin{aligned}\max_{1\leq i\leq N}|A_{i4}| &\leq J^{-1}\max_{1\leq i\leq N}\mu_{\max}\left(J\hat{Q}_{i,\tilde{z}\tilde{z}}\right)\max_{1\leq i\leq N}\left\|\tilde{\pi}_{i}-\pi_{i}^{0}\right\|^{2},\text{ and}\\ \max_{1\leq i\leq N}|A_{i5}| &\leq J^{-1/2}\max_{1\leq i\leq N}\left\|J^{1/2}\hat{Q}_{i,\tilde{z}\tilde{e}}\right\|\max_{1\leq i\leq N}\left\|\tilde{\pi}_{i}-\pi_{i}^{0}\right\|.\end{aligned}$$

Then the results follow from Lemmas A.3(i) and A.4(iii). ■

Proof of Lemma A.6. (i) Noting that $\mathbf{1}\{i\in\hat{G}_{k}\}=\mathbf{1}\{i\in G_{k}^{0}\}+\mathbf{1}\{i\in\hat{G}_{k}\setminus G_{k}^{0}\}-\mathbf{1}\{i\in G_{k}^{0}\setminus\hat{G}_{k}\}$, we have

$$J\left(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)=\frac{J}{N_{k}T}\sum_{i\in\hat{G}_{k}\setminus G_{k}^{0}}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}^{\prime}-\frac{J}{N_{k}T}\sum_{i\in G_{k}^{0}\setminus\hat{G}_{k}}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}^{\prime}\equiv Q_{k,1}-Q_{k,2},\text{ say.}$$

For any $\epsilon>0$, we have $P\left(\|Q_{k,1}\|\geq\epsilon J^{-1/2}\right)\leq P(\hat{F}_{k,NT})\to0$, and $P\left(\|Q_{k,2}\|\geq\epsilon J^{-1/2}\right)\leq P(\hat{E}_{k,NT})\to0$ by Theorem 4.3. Thus $||J(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})||=o_{P}\left(J^{-1/2}\right)$ and (i) follows.

(ii) By the definition of maximum and minimum eigenvalues, we have

$$\begin{aligned}\mu_{\max}\left(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right) &= \sup_{\|\varkappa\|=1}\varkappa^{\prime}\left[\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}+\left(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)\right]\varkappa\leq\mu_{\max}\left(\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)+\left\|\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right\|,\text{ and}\\ \mu_{\min}\left(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right) &= \inf_{\|\varkappa\|=1}\varkappa^{\prime}\left[\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}+\left(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)\right]\varkappa\geq\mu_{\min}\left(\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right)-\left\|\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}-\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}\right\|.\end{aligned}$$

Using arguments analogous to those used in the proof of Lemma A.3(i), we can show that there exist finite positive constants $\underline{c}_{zz}$ and $\bar{c}_{zz}$ such that $P(\underline{c}_{zz}\leq\mu_{\min}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\leq\mu_{\max}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\leq\bar{c}_{zz})\to1$. Then (ii) follows by (i). ■

Proof of Lemma A.7. (i) Noting that $J\hat{Q}_{i,\tilde{z}\tilde{z}}^{(\sigma)}=\frac{J}{T}\sum_{t=1}^{T}\sigma_{i}^{2}\left(X_{it}\right)\tilde{Z}_{it}\tilde{Z}_{it}^{\prime}$, the proof is analogous to that of Lemma A.3(i) and thus omitted. The major change is to replace X_{it} by $\sigma_{i}\left(X_{it}\right)X_{it}$ and apply Assumptions A3(ii)-(iii) in place of Assumptions A1(iii)-(iv). Note that Lemma A.2(i) continues to hold when we replace X_{it} by $\sigma_{i}\left(X_{it}\right)X_{it}$ in the statement.

(ii) Noting that $\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{(\sigma)}=\frac{1}{N_{k}T}\sum_{i\in G_{k}^{0}}\sum_{t=1}^{T}\sigma_{i}^{2}\left(X_{it}\right)\tilde{Z}_{it}\tilde{Z}_{it}^{\prime}$, we can prove the claim by following the arguments used in the proof of Lemma A.3(i) and Assumptions A3(ii)-(iii). ■

Proof of Lemma A.8. (i) First, by the properties of B-splines, and the fact that $\|A\otimes B\|=\|A\|\,\|B\|$ and $\|\mathbf{c}\|=1$,

$$\|b_{\mathbf{c}}\|=\|B\left(v\right)\|\,\|\mathbf{c}\|=\left\{\sum_{j=-d+1}^{J_{0}}B_{j}\left(v\right)^{2}\right\}^{1/2}\leq\left\{\sum_{j=-d+1}^{J_{0}}B_{j}\left(v\right)\right\}^{1/2}=1.$$

Since $B_{j}\left(v\right)=0$ if $v\notin[v_{j},v_{j+d})$, at most $d+1$ of $B_{j}\left(v\right)$'s are nonzero and sum up to one. It follows that $\|b_{\mathbf{c}}\|$ must be bounded from below by a positive constant and $\|b_{\mathbf{c}}\|\asymp1$. Using the inequality $a'Aa \geq \mu_{\min}(A)\, a'a$ repeatedly for conformable vector a and p.s.d. symmetric matrix A and the fact that $\mu_{\min}(A^{-1}) = [\mu_{\max}(A)]^{-1}$ when A is nonsingular, we have by Assumption A3 and Lemmas A.6-A.7

$$\begin{aligned}
S_{\mathbf{c},k}^{2} &= b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{J}{N_kT}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}b_{\mathbf{c}} \\
&\geq \frac{1}{T}\mu_{\min}\left(\frac{J}{N_k}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\right)\left[\mu_{\max}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}\|b_{\mathbf{c}}\|^2 \\
&\geq \min_{i\in G_k^0}\mu_{\min}(V_i)\,\mu_{\min}\left(\frac{J}{N_kT}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i\tilde{Z}_i\right)\left[\mu_{\max}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}\|b_{\mathbf{c}}\|^2 \\
&= \min_{i\in G_k^0}\mu_{\min}(V_i)\,\mu_{\min}\left(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{(\sigma)}\right)\left[\mu_{\max}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}\|b_{\mathbf{c}}\|^2 \\
&\gtrsim \|b_{\mathbf{c}}\|^2 \asymp 1. \qquad\qquad \text{(B.3)}
\end{aligned}$$

(ii) Recall that $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. Let $\tilde{Z}_i = (\tilde{Z}_{i1}, ..., \tilde{Z}_{iT})'$ and $u_i = (u_{i1}, ..., u_{iT})'$. Noting that $\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{u}_{it} = \sum_{t=1}^{T}\tilde{Z}_{it}u_{it} = \tilde{Z}_i'u_i = \tilde{Z}_i'\Sigma_i^{1/2}\varepsilon_i$, we have

$$A_{kNT} = b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}u_i = \sum_{i\in G_k^0}a_i\xi_i,$$

where $a_i = \{\frac{1}{N_kT/J}b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}b_{\mathbf{c}}\}^{1/2}$ and ξ_i's are independent with mean zero and variance one conditional on $\mathcal{X} \equiv \{(X_1, ..., X_N)\}$, (c.f. Lemma A.8 in Huang, Wu, and Zhou 2004). In view of the fact that $E(a_i\xi_i|\mathcal{X}) = 0$, it suffices to prove that $A_{kNT}/\sqrt{\sum_{i\in G_k^0}a_i^2} \xrightarrow{D} N(0,1)$ by verifying the Lindeberg condition:

$$\frac{\max_{i\in G_k^0}a_i^2}{\sum_{i\in G_k^0}a_i^2} = o_p(1). \qquad\qquad \text{(B.4)}$$

Using the inequality $a'Aa \leq \mu_{\max}(A)\, a'a$ repeatedly for conformable vector a and p.s.d. symmetric matrix A and the fact that $\mu_{\max}(A^{-1}) = [\mu_{\min}(A)]^{-1}$ when A is nonsingular, we have by Lemmas A.6-A.7 and Assumptions A3 and A1(vi),

$$\begin{aligned}
\max_{i\in G_k^0}a_i^2 &\leq \frac{J}{N_kT}\max_{i\in G_k^0}\left\{\mu_{\max}\left(\tilde{Z}_i'\Sigma_i^{1/2}V_i\Sigma_i^{1/2}\tilde{Z}_i\right)\right\}b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}b_{\mathbf{c}} \\
&\leq \frac{1}{N_k}\max_{i\in G_k^0}\mu_{\max}(V_i)\max_{i\in G_k^0}\mu_{\max}\left(J\hat{Q}_{i,\tilde{z}\tilde{z}}^{(\sigma)}\right)\left[\mu_{\min}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}\|b_{\mathbf{c}}\|^2 \\
&= \frac{1}{N_k}O_P(\delta_{NT})\,O_P(1)\,O_P(1) = O_P(\delta_{NT}/N) = o_P(1).
\end{aligned}$$

This, in conjunction with (B.3) and the fact that $S_{\mathbf{c},k}^2 = \sum_{i\in G_k^0} a_i^2$, implies that

$$\begin{aligned}
\frac{\max_{i\in G_k^0} a_i^2}{\sum_{i\in G_k^0} a_i^2} &\leq \frac{\frac{1}{N_k}\max_{i\in G_k^0}\mu_{\max}(V_i)\max_{i\in G_k^0}\mu_{\max}\left(J\hat{Q}_{i,\tilde{z}\tilde{z}}^{(\sigma)}\right)\left[\mu_{\min}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}}{\min_{i\in G_k^0}\mu_{\min}(V_i)\,\mu_{\min}\left(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}}^{(\sigma)}\right)\left[\mu_{\max}(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right]^{-2}} \\
&\asymp \frac{1}{N_k}\max_{i\in G_k^0}\mu_{\max}(V_i) = O_P(\delta_{NT}/N) = o_P(1).
\end{aligned}$$

That is, (B.4) is satisfied and the proof of the lemma is complete.

(iii) Let $\zeta_{kNT} = \frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\tilde{Z}_i'\Sigma_i^{1/2}u_i$. By straightforward moment calculation and Chebyshev inequality, $\|\zeta_{kNT}\| = O_P\left(J^{1/2}\right)$. Noting that $\hat{A}_{kNT}-A_{kNT} = -b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}[J(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} -\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})](J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\zeta_{kNT}$ and $\|b_{\mathbf{c}}\| \leq 1$, we have

$$\begin{aligned}
\left|\hat{A}_{kNT} - A_{kNT}\right| &\leq \left|b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\left[J(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} - \bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right](J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\zeta_{kNT}\right| \\
&\leq \left\|(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\right\|_{\text{sp}}\left\|J(\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}} - \bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})\right\|\left\|(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\right\|_{\text{sp}}\|\zeta_{kNT}\| \\
&= O_P(1)\,o_P\left(J^{-1/2}\right)O_P(1)\,O_P\left(J^{1/2}\right) = o_P(1).
\end{aligned}$$

It follows that $(\hat{A}_{kNT} - A_{kNT})/S_{\mathbf{c},k} = o_P(1)$ by (i). ■

Proof of Lemma A.9. Let $\hat{C}_{kNT} = \bar{b}_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\xi_{kNT}$ where $\xi_{kNT} = \frac{1}{\sqrt{N_kT/J}}\sum_{i\in G_k^0}\sum_{t=1}^T\tilde{Z}_{it}\tilde{\eta}_{it}$. By Assumption A1 and Markov inequality,

$$\begin{aligned}
\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\eta_{it}^2 &= \frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\left\{\left[\beta_i^0(t/T) - \pi_i^0B(t/T)\right]'X_{it}\right\}^2 \\
&\leq \vartheta_{NT}^2\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\|X_{it}\|^2 = O_P\left(\vartheta_{NT}^2\right).
\end{aligned}$$

Next, let $\boldsymbol{\varpi}^* = (J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}b_{\mathbf{c}}$. Then by Lemma A.2,

$$\begin{aligned}
\frac{J}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\left|b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\tilde{Z}_{it}\right|^2 &= \frac{J}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\boldsymbol{\varpi}^{*\prime}\tilde{Z}_{it}\tilde{Z}_{it}'\boldsymbol{\varpi}^* \\
&= \frac{J}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T g_{\boldsymbol{\varpi}^*}(t/T)'X_{it}X_{it}'g_{\boldsymbol{\varpi}^*}(t/T) \\
&= \frac{J}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T g_{\boldsymbol{\varpi}^*}(t/T)'E\left[X_{it}X_{it}'\right]g_{\boldsymbol{\varpi}^*}(t/T)\{1+o_P(1)\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\bar{c}_{xx}J}{T}\sum_{t=1}^{T} g_{\boldsymbol{\varpi}^*}(t/T)' g_{\boldsymbol{\varpi}^*}(t/T) = \frac{J\bar{c}_{xx}}{T}\sum_{t=1}^{T} \boldsymbol{\varpi}^{*\prime} B(t/T) B(t/T)' \boldsymbol{\varpi}^* \\
&\asymp J\boldsymbol{\varpi}^{*\prime}\int_0^1 B(v) B(v)' dv \boldsymbol{\varpi}^* \asymp \|\boldsymbol{\varpi}^*\| = \|\boldsymbol{\varpi}^*\|_{\text{sp}} \leq \left\|(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\right\|_{\text{sp}} = O_P(1).
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
\left|\hat{C}_{kNT}\right| &\leq \frac{1}{\sqrt{N_k T/J}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\left|b_{\mathbf{c}}'(J\hat{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\tilde{Z}_{it}\right| |\eta_{it}| \\
&\leq \sqrt{N_k T}\left\{\frac{J}{N_k T}\sum_{i\in G_k^0}\sum_{t=1}^{T}\left|b_{\mathbf{c}}'(J\bar{\mathbb{Q}}_{k,\tilde{z}\tilde{z}})^{-1}\tilde{Z}_{it}\right|^2\right\}^{1/2}\left\{\frac{1}{N_k T}\sum_{i\in G_k^0}\sum_{t=1}^{T}\eta_{it}^2\right\}^{1/2} \\
&\leq \sqrt{N_k T} O_P(1) O_P(\vartheta_{NT}) = o_P(1).
\end{aligned}
$$

This completes the proof of the lemma. ■

Proof of Lemma A.10. The proof is analogous to that of Lemma A.1 in Su, Shi, and Phillips (2016). When $K \geq K_0$, we can follow the proof of Theorem 4.1 and show that $\|\tilde{\pi}_i - \pi_i^0\| = O_P(J^{-\gamma+1/2} + JT^{-1/2} + \lambda J^{(K+1)/2})$ for each i and $\frac{1}{N}\sum_{i=1}^N \prod_{k=1}^K \Pi_{k=1}^K \|\pi_i^0 - \hat{\omega}_k\| = O_P(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2})$. Noting that π_i^0, $i = 1, ..., N$, only take K_0 distinct values, the latter implies that the collection $\{\hat{\omega}_k, k = 1, ..., K\}$ contains at least K_0 distinct vectors, say, $\hat{\omega}_{(1)}, ..., \hat{\omega}_{(K_0)}$, such that $\hat{\omega}_{(k)} - \alpha_k^0 = J^{-(K-1)/2} O_P(J^{-\gamma+K/2} + J^{(K+1)/2}T^{-1/2}) = O_P(J^{-\gamma+1/2} + JT^{-1/2})$ for $k = 1, ..., K_0$. For notational simplicity, we rename the other vectors in the above collection as $\hat{\omega}_{(K_0+1)}, ..., \hat{\omega}_{(K)}$. As before, we classify $i \in \hat{G}_k(K,\lambda)$ if $\left\|\hat{\pi}_i - \hat{\omega}_{(k)}\right\| = 0$ for $k = 1, ..., K$, and $i \in \hat{G}_0(K,\lambda)$ otherwise. Using arguments as used in the proof of Theorem 4.3, we can show that

$$
\sum_{i\in G_k^0} P\left(\hat{E}_{k,NT,i}\right) = o(1) \text{ for } k = 1, ..., K_0 \text{ and } \sum_{i\in \hat{G}_k(K,\lambda)} P\left(\hat{F}_{k,NT,i}\right) = o(1) \text{ for } k = 1, ..., K_0.
$$

The first part implies that $\sum_{i=1}^N P(i \in \hat{G}_0(K,\lambda) \cup \hat{G}_{K_0+1}(K,\lambda) \cup ... \cup \hat{G}_K(K,\lambda)) = o(1)$.

Let $\hat{u}_{it}(k) = \tilde{Y}_{it} - \tilde{Z}_{it}'\text{vec}\left(\hat{\omega}_{\hat{G}_k(K,\lambda)}\right)$. Observe that

$$
\hat{u}_{it}(k) = \tilde{Y}_{it} - \tilde{X}_{it}'\hat{\alpha}_{\hat{G}_k(K,\lambda)}(t/T) = \tilde{e}_{it} - \tilde{X}_{it}'\left[\hat{\alpha}_{\hat{G}_k(K,\lambda)}(t/T) - \alpha_k^0(t/T)\right]. \tag{B.5}
$$

Using the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\} - \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\}$, we have $\hat{\sigma}^2_{\hat{G}(K,\lambda)} = \frac{1}{NT}\sum_{k=1}^K \sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^T [\hat{u}_{it}(k)]^2 = D_{1NT} + D_{2NT} - D_{3NT} + D_{4NT}$, where

$$
\begin{aligned}
D_{1NT} &= \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\sum_{t=1}^{T}[\hat{u}_{it}(k)]^2, \; D_{2NT} = \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K,\lambda)\backslash G_k^0}\sum_{t=1}^{T}[\hat{u}_{it}(k)]^2, \\
D_{3NT} &= \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0\backslash\hat{G}_k(K,\lambda)}\sum_{t=1}^{T}[\hat{u}_{it}(k)]^2, \text{ and } D_{4NT} = \frac{1}{NT}\sum_{k=K_0+1}^{K}\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^{T}[\hat{u}_{it}(k)]^2.
\end{aligned}
$$

Following the proof of Theorem 4.4, we can show that $\hat{\alpha}_{\hat{G}_k(K,\lambda)}(t/T)-\alpha_k^0(t/T)=O_P((NT/J)^{-1/2})$ for $k=1,...,K_0$. Using this and (B.5), we can readily show that $D_{1NT}=\bar{\sigma}_{G^0}^2+O_P((NT/J)^{-1})$. For D_{2NT}, D_{3NT}, and D_{4NT}, we have that for any $\epsilon>0$, $P(D_{2NT}\geq(NT/J)^{-1}\epsilon)\leq\sum_{i=1}^{K_0}P(\hat{F}_{k,NT})\to0$, $P(D_{3NT}\geq(NT/J)^{-1}\epsilon)\leq\sum_{i=1}^{K_0}P(\hat{E}_{k,NT})\to0$, and $P(D_{4NT}\geq(NT/J)^{-1}\epsilon)\leq\sum_{i=1}^{N}P(i\in\cup_{K_0+1\leq k\leq K}\hat{G}_k(K,\lambda))\to0$. It follows that $\hat{\sigma}_{\hat{G}(K,\lambda)}^2=\bar{\sigma}_{G^0}^2+O_P((NT/J)^{-1})$ for all $K_0\leq K\leq K_{\max}$.■

C Numerical Algorithm

Here, we follow Su, Shi, and Phillips (2014) and propose an iterative numerical algorithm to obtain the estimates $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\omega}}$. The iterative algorithm goes as follows:

1. Start with arbitrary initial values $\hat{\boldsymbol{\omega}}^{(0)}=(\text{vec}(\hat{\omega}_1^{(0)})',\text{vec}(\hat{\omega}_2^{(0)})',...,\text{vec}(\hat{\omega}_K^{(0)})')'$ and $\hat{\boldsymbol{\pi}}^{(0)}=(\text{vec}(\hat{\pi}_1^{(0)})',\text{vec}(\hat{\pi}_2^{(0)})',...,\text{vec}(\hat{\pi}_N^{(0)})')'$ such that $\sum_{i=1}^{N}\|\hat{\pi}_i^{(0)}-\hat{\omega}_k^{(0)}\|\neq0$ for each $k=2,...,K$. For example, one can simply let $\hat{\pi}_i^{(0)}=\tilde{\pi}_i$ for $i=1,...,N$ and set $\hat{\omega}_k^{(0)}$'s to be either the zero vector or the average of $\hat{\pi}_i^{(0)}$'s.

2. In step $r\geq1$, given $\hat{\boldsymbol{\omega}}^{(r-1)}=(\text{vec}(\hat{\omega}_1^{(r-1)})',...,\text{vec}(\hat{\omega}_K^{(r-1)})')'$ and $\hat{\boldsymbol{\pi}}^{(r-1)}=(\text{vec}(\hat{\pi}_1^{(r-1)})',...,\text{vec}(\hat{\pi}_N^{(r-1)})')'$, we first choose $(\boldsymbol{\pi},\omega_1)$ to minimize

$$Q_{NT}^{(r,1)}(\boldsymbol{\pi},\omega_1)=Q_{1,NT}(\boldsymbol{\pi})+\frac{\lambda}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\left\|\tilde{V}_i\text{vec}(\pi_i-\omega_1)\right\|\times\prod_{k\neq1}^{K}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i^{(r-1)}-\hat{\omega}_k^{(r-1)})\right\|,$$

and obtain the updated estimate $(\hat{\boldsymbol{\pi}}^{(r,1)},\hat{\omega}_1^{(r)})$ of $(\boldsymbol{\pi},\omega_1)$. Then, we choose $(\boldsymbol{\pi},\omega_2)$ to minimize

$$Q_{NT}^{(r,2)}(\boldsymbol{\pi},\omega_2)=Q_{1,NT}(\boldsymbol{\pi})+\frac{\lambda}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\left\|\tilde{V}_i\text{vec}(\pi_i-\omega_2)\right\|\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i^{(r,1)}-\hat{\omega}_1^{(r)})\right\|$$
$$\times\prod_{k\neq1,2}^{K}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i^{(r-1)}-\hat{\omega}_k^{(r-1)})\right\|$$

and obtain the updated estimate $(\hat{\boldsymbol{\pi}}^{(r,2)},\hat{\omega}_2^{(r)})$ of $(\boldsymbol{\pi},\omega_2)$. Repeat this procedure until $(\boldsymbol{\pi},\omega_K)$ is chosen to minimize

$$Q_{NT}^{(r,K)}(\boldsymbol{\pi},\omega_2)=Q_{1,NT}(\boldsymbol{\pi})+\frac{\lambda}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^{2-K}\left\|\tilde{V}_i\text{vec}(\pi_i-\omega_K)\right\|\prod_{k=1}^{K-1}\left\|\tilde{V}_i\text{vec}(\hat{\pi}_i^{(r)}-\hat{\omega}_k^{(r)})\right\|,$$

to obtain the updated estimate $(\hat{\boldsymbol{\pi}}^{(r,K)},\hat{\omega}_K^{(r)})$ of $(\boldsymbol{\pi},\omega_K)$. Let $\hat{\boldsymbol{\pi}}^{(r)}=\hat{\boldsymbol{\pi}}^{(r,K)}$ and $\hat{\boldsymbol{\omega}}^{(r)}=(\text{vec}(\hat{\omega}_1^{(r)})',...,\text{vec}(\hat{\omega}_K^{(r)})')'$.

3. Repeat step 2 until a convergence criterion is reach, e.g., when

$$\frac{\sum_{i=1}^{N}\left\|\pi_i^{(r)}-\pi_i^{(r-1)}\right\|^2}{\sum_{i=1}^{N}\left\|\pi_i^{(r-1)}\right\|^2+0.0001}<\epsilon_{tol}, \text{ and } \frac{\sum_{k=1}^{K}\left\|\omega_k^{(r)}-\omega_k^{(r-1)}\right\|^2}{\sum_{k=1}^{N}\left\|\omega_k^{(r-1)}\right\|^2+0.0001}<\epsilon_{tol}$$

where ϵ_{tol} is some pre-specified tolerance level (e.g., 0.0001). Define the final (say R) iterative estimate of $\boldsymbol{\omega}$ which satisfies the convergence criterion as $\hat{\boldsymbol{\omega}} = (\text{vec}(\hat{\omega}_1^{(R)})', ..., \text{vec}(\hat{\omega}_K^{(R)})')'$ and the final estimate of $\boldsymbol{\pi}$ as $\hat{\boldsymbol{\pi}} = (\text{vec}(\hat{\pi}_1)', ..., \text{vec}(\hat{\pi}_N)')'$, where

$$\begin{aligned}
\hat{\pi}_i = & \sum_{k=1}^{K}\hat{\omega}_k^{(R)}\mathbf{1}\left\{\hat{\pi}_i^{(R,l)}=\hat{\omega}_k^{(R)} \text{ for some } l=1,...,K\right\} \\
& +\hat{\pi}_i^{(R,K)}\left[1-\sum_{k=1}^{K}\mathbf{1}\left\{\hat{\pi}_i^{(R,l)}=\hat{\omega}_k^{(R)} \text{ for some } l=1,...,K\right\}\right]
\end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. That is, individual i is classified to group $\hat{G}_k$ if $\pi_i^{(R,l)} = \hat{\omega}_k^{(R)}$ for some $l = 1, 2, ..., K$; otherwise, it is left unclassified and we can define $\hat{\pi}_i$ as $\hat{\pi}_i^{(R,K)}$.

We note that the objective function $Q_{NT}^{(r,k)}(\boldsymbol{\pi}, \omega_k)$ is convex in $(\boldsymbol{\pi}, \omega_k)$ in each substep k of the rth iteration. Therefore, the above iterative algorithm can be implemented with rapid convergence in practice.

D Further Details on the Empirical Application

D.0.1 Estimation Results

When we fix the number of groups to be four, our C-Lasso method with $c_\lambda = 1$ classifies the 91 economies into the following four groups:

- Group 1 (16 Economies): Belize, Botswana, China, Egypt, India, Indonesia, Japan, Korea, Luxembourg, Malaysia, Portugal, Puerto Rice, Saint Vincent and the Grenadines, Seychelles, Singapore, Sri Lanka;
- Group 2 (31 Economies): Australia, Austria, Barbados, Belgium, Bermuda, Brazil, Canada, Chile, Colombia, Costa Rica, Denmark, Dominican, Finland, France, Greece, Hungary, Iceland, Israel, Italy, Lesotho, Morocco, Netherlands, Norway, Pakistan, Paraguay, Spain, Sweden, Turkey, United Kingdom, United States, World;
- Group 3 (26 Economies): Algeria, Bahamas, Bangladesh, Benin, Burkina Faso, Cameroon, Republic of the Congo, Ecuador, Fiji, Gabon, Guatemala, Guyana, Honduras, Kenya, Malawi, Mexico, Nepal, Papua New Guinea, Panama, Peru, Philippines, Rwanda, South Africa, Sudan, Trinidad and Tobago, Uruguay;

- Group 4 (18 Economies): Burundi, Bolivia, Bolivarian Republic of Venezuela, Central African Republic, Chad, Republic of Cote D'Ivoire, Democratic Republic of Congo, Ghana, Madagascar, Mauritania, Niger, Nigeria, Nicaragua, Senegal, Sierra Leone, Togo, Zambia, Zimbabwe.

Figure 4 in the paper depicts the estimated trends for the four estimated groups and Figure 5 reports the realization of GDP per capita (logarithm and demeaned) and the trend for each group. As shown in the figures, the first group contains 16 economies with fastest growth rate of GDP per capita. Half of them are emerging Asian countries and the other eight economies are from Africa (3), Europe (2), North America (2) and South America (1). The trend of the second group is quite similar to the path of the whole world's GDP per capita. Most of the countries in the second group are developed countries with 20 out of them being OECD countries. The third groups are countries with lower but still positive growth rate. Most of these countries belong to Latin America and Africa, which either have low GDP per capita or suffer from the "middle income trap". The trend of the fourth group is oscillating about the zero line, which means that the improvement of the GDP per capita is negligible during the last half century. 15 out of 18 countries in group 4 are from Africa, which may be the poorest countries in the world, while the other three are from Latin America and suffer severely from the "middle income trap".

Our results have some implications for economic modeling and testing. First, although the estimated trend of the first group's GDP per capita appear to be approximately linear, the other three estimated trends appear to be nonlinear. Thus, when one tests panel unit root against trend stationary processes, the setting of linear trends may suffer from model misspecification problem and induces misleading results. Second, most of the developing countries in Asia, which are generally classified to the first group, provide justifications for the Neoclassical growth models' implication of world-wide convergence, as the first group grows faster than the second group, containing most of developed countries. However, this is not the case for Latin America and Africa, which generally belong to the third and fourth groups. Both the level and the growth rate of GDP per capita for countries from these two continents are lower than those for the developed countries. As a result, the gap between these two continents and the developed countries is widened, i.e., it shows economic divergence rather convergence. Third, although our data-driven classification exhibits some geographic features, it can not be obtained based on geographic location. For example, such Asian countries as Philippines and Bangladesh are classified into the third group. In addition, although most of the OECD countries are classified into the second group, there are still some countries classified into the first (e.g., Japan, Korea, Luxembourg, and Portugal) and the third (e.g., Mexico) groups. As a result, a classification based on some external criterion such as continental location is inevitably misleading.

REFERENCE

Bernstein, D. S. (2005), *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective.* Springer, New York.

Huang, J.Z. and Shen H. (2004), "Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach," *Scandinavian Journal of Statistics*, 31, 515-534.

Huang, J.Z., Wu C. O., and Zhou L. (2004), "Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data," *Statistica Sinica*, 14, 763-788.

Merlevède, F., Peligrad, M., and Rio, E. (2009), "Bernstein Inequality and Moderate Deviations under Strong Mixing Conditions," IMS collections. *High Dimensional Probability V.*, 273-292.

Qian, J. and Su, L. (2016), "Shrinkage Estimation of Regression Models with Multiple Structural Changes," *Econometric Theory*, 32, 1376-1433.

Su, L., Shi, Z. and Phillips, P. C. B. (2016), "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215-2264.